

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2001, 13:03:12 ; Search time 88.53 Seconds  
(without alignments)  
2949.726 Million cell updates/sec

Title: US-09-508-967-1

Perfect score: 12100  
Sequence: 1 MATSGSGGTQEDAKHVLD.....VNKKKEFEERYPISDIWT 2228

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_15:\*

- 1: sp.archaea:\*
- 2: sp.bacteria:\*
- 3: sp.fungi:\*
- 4: sp.human:\*
- 5: sp.invertebrate:\*
- 6: sp.mammal:\*
- 7: sp.mhc:\*
- 8: sp.organelle:\*
- 9: sp.phage:\*
- 10: sp.plant:\*
- 11: sp rodent:\*
- 12: sp.unclassified:\*
- 13: sp.vertebrate:\*
- 14: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12065	99.7	2228	5	060991 plasmodium
2	5216.5	43.1	2277	5	090095 plasmodium
3	5074	41.9	2197	5	096296 plasmodium
4	4901	40.5	3078	5	026031 plasmodium
5	4817	39.8	2182	5	026034 plasmodium
6	4809.5	39.7	2710	5	09X2B8 plasmodium
7	4711	38.9	2664	5	026033 plasmodium
8	4648.5	38.4	2209	5	090066 plasmodium
9	4493.5	37.1	2163	5	09NFB6 plasmodium
10	4341.5	35.9	2169	5	097312 plasmodium
11	4302.5	35.6	2209	5	097324 plasmodium
12	4084	33.8	2647	5	090580 plasmodium
13	4073	33.7	2042	5	025766 plasmodium
14	3989	33.0	3006	5	026032 plasmodium
15	3937	32.5	2924	5	025733 plasmodium
16	3297.5	27.3	1711	5	096108 plasmodium
17	3192	26.4	2706	5	015870 plasmodium
18	2959	24.5	1729	5	025734 plasmodium
19	2863	23.7	2212	5	094657 plasmodium

20	2668	22.0	3026	5	026030 plasmodium
21	2518	20.8	2135	5	061077 plasmodium
22	1784	14.7	438	5	09Y1N7 plasmodium
23	1771.5	14.6	435	5	09Y1N6 plasmodium
24	1758.5	14.5	1327	5	09NFB4 plasmodium
25	1681.5	13.9	3542	5	09U5M2 plasmodium
26	1629	13.5	1685	5	09U4A2 plasmodium
27	1356.5	11.2	457	5	09NC63 plasmodium
28	1319.5	10.9	431	5	096294 plasmodium
29	782	6.5	440	5	096110 plasmodium
30	777	6.4	259	5	09U7K1 plasmodium
31	728	6.0	255	5	09U7H6 plasmodium
32	724.5	6.0	247	5	09U7I7 plasmodium
33	720	6.0	480	5	09NAV6 plasmodium
34	719	5.9	254	5	09U7I8 plasmodium
35	708.5	5.9	258	5	09U7I2 plasmodium
36	702	5.8	921	5	025989 plasmodium
37	695	5.7	260	5	09U7I5 plasmodium
38	689.5	5.7	251	5	09U7H5 plasmodium
39	677.5	5.6	254	5	09U7J2 plasmodium
40	661.5	5.5	247	5	09U7I9 plasmodium
41	644.5	5.3	267	5	061076 plasmodium
42	641	5.3	246	5	09U7J6 plasmodium
43	622	5.1	298	5	061061 plasmodium
44	615.5	5.1	245	5	09U7I4 plasmodium
45	614.5	5.1	236	5	09U7H7 plasmodium

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	2228 AA.
060991	060991			
AC	01-AUG-1998 (TREMBLrel. 07, Created)			
DT	01-AUG-1998 (TREMBLrel. 07, last sequence update)			
DT	01-JUN-2000 (TREMBLrel. 14, last annotation update)			
DE	ERYTHROCYTE MEMBRANE PROTEIN 1.			
GN	FCR3S1.2-VARI.			
OS	Plasmodium falciparum.			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5833;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98080592; PubMed=9419207;			
RA	Chen Q., Barragan A., Fernandez V., Sundstrom A., Schlichtherle M.,			
RA	Sahlen A., Carlson J., Datta S., Wahlgren M.;			
RT	Identification of Plasmodium falciparum erythrocyte membrane protein			
RT	1 (PFEMP1) as the resetting ligand of the malaria parasite P.			
RT	falciparum."			
RT	J. Exp. Med. 187:15-23(1998).			
DR	EMBL; AF003473; AAC05730.1; "			
RL	SEQUENCE 2228 AA; 252811 MW; 5D8C8E9BFA22DC8B CRC4;			
SQ				
Query Match	99.7%;	Score 12065;	DB 5;	Length 2228;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 2223;	Conservative	0;	Mismatches 5;	Indels 0;
Gaps	0;			
QY	1	MATSGSGGTQEDAKHVLD	FEFGOKVHDEHGAKYVSELGSLA	IIILETAFYVS 60
DB	1	MATSGSGGTQEDAKHVLD	FEFGOKVHDEHGAKYVSELGSLA	IIILETAFYVS 60
QY	61	MOTESYTELIEANSKRNPCKDKGNDVDRFSVKEQAGYDNKKMKCSIGMTCAFFRRLL 120		
DB	61	MOTESYTELIEANSKRNPCKDKGNDVDRFSVKEQAGYDNKKMKCSIGMTCAFFRRLL 120		
QY	121	LCNKNFPNNNSNDSSRAKHDLAEVCMARKYEGESIKTHYPYDSKYPISDPFMCMTLAR 180		
DB	121	LCNKNFPNNNSNDSSRAKHDLAEVCMARKYEGESIKTHYPYDSKYPISDPFMCMTLAR 180		
QY	181	SFADIGDITRGNDLYGNKKKKKNGKTEERKLEQKLKLFKKIHNTLQKFAKRYND 240		

```
|||||
Db 181 SFAIGDIIIRDRDLYLKGKKKKONGKETERERKLEQKLEIKFIKKIHDNKLKDEQAQRYNGD 240
Oy 241 EDPNFKYLRDMWMTANRFTVGMATCSELDNSSYFRATCNDTGGSPQOTHNKCDCDK 300
Db 241 EDPNFKYLRDMWMTANRFTVGMATCSELDNSSYFRATCNDTGGSPQOTHNKCDCDK 300
Oy 301 GANAGKPRAGDGDVYTYVFPYQYLRWFEWNAEDFCRRKKKKLEMLEKQCRGKDKSD 360
Db 301 GANAGKPRAGDGDVYTYVFPYQYLRWFEWNAEDFCRRKKKKLEMLEKQCRGKDKSD 360
Oy 361 EYRCSRNGYCEQOTISKKGKVRMGKCTDCFPACGSTEYNNIDNOROKPDKOKYTKIS 420
Db 361 EYRCSRNGYCEQOTISKKGKVRMGKCTDCFPACGSTEYNNIDNOROKPDKOKYTKIS 420
Oy 421 DGGRRKKRVAAGTTRKYBEGEKSFEYKLLKNDGYTDAFLGLNNKKAOKDTTGGKTIK 480
Db 421 DGGRRKKRVAAGTTRKYBEGEKSFEYKLLKNDGYTDAFLGLNNKKAOKDTTGGKTIK 480
Oy 481 EVNSGGVYVGGSGGTSAGSTNDENKGTFRSEYQPCPCGYOHKGNOMERTKYK 540
Db 481 EVNSGGVYVGGSGGTSAGSTNDENKGTFRSEYQPCPCGYOHKGNOMERTKYK 540
Oy 541 MRMSKLYPIGKNVLLKSLKYVKDMILKKNKEFCLTQNSSDGSYGVYTTGASG 600
Db 541 MRMSKLYPIGKNVLLKSLKYVKDMILKKNKEFCLTQNSSDGSYGVYTTGASG 600
Oy 601 SEKKELDEMKCYHNEQKVNVOGEVEDDELKAGGLCILPPKKNKKNVSEAKSGNN 660
Db 601 SEKKELDEMKCYHNEQKVNVOGEVEDDELKAGGLCILPPKKNKKESEKSSNN 660
Oy 661 HADIOKTFHDFEYVVAHMLKDSIHWRTKRLKCSIDGKTKCRNGKCKDCCEKWNQ 720
Db 661 HADIOKTFHDFEYVVAHMLKDSIHWRTKRLKCSIDGKTKCRNGKCKDCCEKWNQ 720
Oy 721 KETEMKPKTDHFKTOEGTPEGYFTTLELILKLOFLKEDTEBNTBNSIDAEBEKLQ 780
Db 721 KETEMKPKTDHFKTOEGTPEGYFTTLELILKLOFLKEDTEBNTBNSIDAEBEKLQ 780
Oy 781 KILLENENNLAVNAGTEOKTLMDKLNLHNLNATKCDPLPEEDKSRRSADPSDI 840
Db 781 KILLENENNLAVNAGTEOKTLMDKLNLHNLNATKCDPLPEEDKSRRSADPSDI 840
Oy 841 FIPPEEKEDDENDEDEYRDEDETAKEETTESATDTTSLDVCPIVGKVLTKDNESL 900
Db 841 FIPPEEKEDDENDEDEYRDEDETAKEETTESATDTTSLDVCPIVGKVLTKDNESL 900
Oy 901 QDASCLKYGKNNRGLGRCVTPSGEPTSSDKNGAICYPPRRRLYIKIYDMATKTESP 960
Db 901 QDASCLKYGKNNRGLGRCVTPSGEPTSSDKNGAICYPPRRRLYIKIYDMATKTESP 960
Oy 961 QASGEASSTGSTTPPSKFAALLKAFVESAIETFFLMHRYKEBKRAVAOEGHGLPR 1020
Db 961 QASGEASSTGSTTPPSKFAALLKAFVESAIETFFLMHRYKEBKRAVAOEGHGLPR 1020
Oy 1021 VEESPEYDPEDLKEGKIPDGLRQMFYTLGDYRDILFSSGNDTTSYKSDTPSSNDNL 1080
Db 1021 VEESPEYDPEDLKEGKIPDGLRQMFYTLGDYRDILFSSGNDTTSYKSDTPSSNDNL 1080
Oy 1081 KNIVLLASGSTEOREKKNKKETKNFRKCTERSAPLVSHPQTWENNKKYIWHGAVC 1140
Db 1081 KNIVLLASGSTEOREKKNKKETKNFRKCTERSAPLVSHPQTWENNKKYIWHGAVC 1140
Oy 1141 ALTSKDKIAKVEKPKOKIEPENLMDPEANKPKPOYQYNNVLDENSGTSPRTTOQA 1200
Db 1141 ALTSKDKIAKVEKPKOKIEPENLMDPEANKPKPOYQYNNVLDENSGTSPRTTOQA 1200
Oy 1201 SSDNTPPTLTHFVAPTYFRWFEEMGESFCRERKKRLKQIVDKVENGDCVGRCSGDEA 1260
Db 1201 SSDNTPPTLTHFVAPTYFRWFEEMGESFCRERKKRLKQIVDKVENGDCVGRCSGDEA 1260
Oy 1261 CDSISTHDYSTVPSPNCBCKHCSSTRKWTERRKIEFHKSNAVGOOKDATRNNGNTF 1320
|||||

|||||
Db 1261 CDSISTHDYSTVPSPNCBCKHCSSTRKWTERRKIEFHKSNAVGOOKDATRNNGNTF 1320
Oy 1321 DKEPCKLETMPDAKFLERLKNGPCNTKKEYGGDIDFEKDSOTFOHTEVCGCPKFKT 1380
Db 1321 DKEPCKLETMPDAKFLERLKNGPCNTKKEYGGDIDFEKDSOTFOHTEVCGCPKFKT 1380
Oy 1381 NCONGNCVSGJLNGCDBKSIDAKELAKMRSSTDDVVRVSDNDTTFEGDLDKACQH 1440
Db 1381 NCONGNCVSGJLNGCDBKSIDAKELAKMRSSTDDVVRVSDNDTTFEGDLDKACQH 1440
Oy 1441 ANIFKGIKDVWKGYCVGVICQOTNINERTDKEYIQTALKRWVENLEBYNKIND 1500
Db 1441 ANIFKGIKDVWKGYCVGVICQOTNINERTDKEYIQTALKRWVENLEBYNKIND 1500
Oy 1501 KISHCIRKGEBSKCIINGCEKSKCLEKIEKIAEMENIKRFPDOYENKQOPYNYKSI 1560
Db 1501 KISHCIRKGEBSKCIINGCEKSKCLEKIEKIAEMENIKRFPDOYENKQOPYNYKSI 1560
Oy 1561 LEBELPKIAVNDODNVIKLCVFENSKGCTLISNTONKENDADICMLKKGVAKACQ 1620
Db 1561 LEBELPKIAVNDODNVIKLCVFENSKGCTLISNTONKENDADICMLKKGVAKACQ 1620
Oy 1621 KPSEKQSDCKEPPPLPDEEONPEBNTLEPPKCPPTTOPPEKGETCGNKEKKDEK 1680
Db 1621 KPSEKQSDCKEPPPLPDEEONPEBNTLEPPKCPPTTOPPEKGETCGNKEKKDEK 1680
Oy 1681 KESEEPKKEESGPAEAPAPTAESEETNFPBPBGPAAPSTPAFPPTPPPLRP 1740
Db 1681 KESEEPKKEESGPAEAPAPTAESEETNFPBPBGPAAPSTPAFPPTPPPLRP 1740
Oy 1741 QADEPFDSTILQTTIPFGVALALGSIAPFLKKTAKASVGNLFQILOPKSDYDIPPLKS 1800
Db 1741 QADEPFDSTILQTTIPFGVALALGSIAPFLKKTAKASVGNLFQILOPKSDYDIPPLKS 1800
Oy 1801 SNRYIPIYSDRYKRGKTYIYMGSDDEKYYAFMSOTTDVTSSESEYEELDINDIYVPSPK 1860
Db 1801 SNRYIPIYSDRYKRGKTYIYMGSDDEKYYAFMSOTTDVTSSESEYEELDINDIYVPSPK 1860
Oy 1861 YKTLIEVYLEPSSGNNTTASGKNTPSDPTFNDIONGDIPSSKITTDEMMOLKKEFISNMLON 1920
Db 1861 YKTLIEVYLEPSSGNNTTASGKNTPSDPTFNDIONGDIPSSKITTDEMMOLKKEFISNMLON 1920
Oy 1921 QPNVNDYTSGNSSTNTNTTTSRHVNDNNTTMSHDNNEMLLPSIHGMLYSGEE 1980
Db 1921 QPNVNDYTSGNSSTNTNTTTSRHVNDNNTTMSHDNNEMLLPSIHGMLYSGEE 1980
Oy 1981 YSYVNNVNSMNDIPINRDNVYSGIDILINDLSGKRPIDYDEVLARKENELFGTEWTK 2040
Db 1981 YSYVNNVNSMNDIPINRDNVYSGIDILINDLSGKRPIDYDEVLARKENELFGTEWTK 2040
Oy 2041 RSTQONVAKTTPNSPIHNOLELFRKWLDRHRDMCEKMKNEEDILNKLEENKKNINSNG 2100
Db 2041 RSTQONVAKTTPNSPIHNOLELFRKWLDRHRDMCEKMKNEEDILNKLEENKKNINSNG 2100
Oy 2101 KTYNSDNKPSHNHVLNLDVSIQIDMDNPKTKNEITNNDTNDKSTWTDILDDLEKYNDPY 2160
Db 2101 KTYNSDNKPSHNHVLNLDVSIQIDMDNPKTKNEITNNDTNDKSTWTDILDDLEKYNDPY 2160
Oy 2161 YYDFYEDDIYHDVDEKSSMDIYVDHNNTYSNMNVPPFMHLENNIYNNKKIEFEERY 2220
Db 2161 YYDFYEDDIYHDVDEKSSMDIYVDHNNTYSNMNVPPFMHLENNIYNNKKIEFEERY 2220
Oy 2221 PISDIWNI 2228
Db 2221 PISDIWNI 2228

RESULT 2
09U0G5 PRELIMINARY; PRT; 2277 AA.
AC 09U0G5:
DT 01-MAY-2000 (TIREMBLrel. 13, Created)
DT 01-MAY-2000 (TIREMBLrel. 13, last sequence update)
```

01-MAY-2000 (TREMREL.13, Last annotation update)  
DE VAR  
GN MALAP2.58.  
OS Plasmodium falciparum.  
OC Eularyota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RA Devlin K., Bowman S., Churcher C., Harris B., Harris D., Lawson D.,  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL035475; CAB62899.1; -;  
SQ SEQUENCE 2277 AA; 256122 MW; E4262CCA69DDEF93 CRC64;

Query Match 43.1%; Score 5216.5; DB 5; Length 2277;  
Best Local Similarity 48.4%; Pred. No. 2.3e-276;  
Matches 1160; Conservative 272; Mismatches 643; Indels 323; Gaps 78;

13 EDAAIVLDFGQKVIDEVHGEAKNT-VSEFLKGSLSLSTLGE--TAFYVKSMTQESKYTE 69  
14 KDAKDLDMIGKDVHDQVEKEAAGSGSELKGLSLAKSGVELAFAFPCGLIKDKGDE 73  
70 LEANSKRNPKCKDKGNDVDRESVKEAGYDNKKMKKCSNGM-----TCAPRRHLICN 123  
74 LAGDSGERHPGCGNTTGKEDVDRESVKQAEYDNKKKCSYSGNGVDGACAPRRLEICN 133  
124 KNFPMNNSDSKAKHDLAEYCAKAKYEGESIKTHYPKDYDSKPGSDPFCMTMLARSPA 183  
134 KMEKMKGR--TSTKHDLLDVCMANAYQAOLIRYHDHQLNBSQ--ICTVLARSPA 189  
184 DIGDIIRGRDLYGNKKKKKQNGEKTEREKLQKLEIFKKIHDLKD--KEAQKRYNGDE 241  
190 DIGDIIRGRDLYGNKKKKKKT--ETERDQLESKLIKIFGDIYNEILNGRNGYKDYODDN 247  
242 DNFFKLRDMWTANREYWGAMTSCKELDNSSYPATGNDP--GGP--SOTINCKCDDND 299  
248 GGNFOLRDMWTANRATYAKAITCADTGN--ATFRPTSDSDGKSFQANDCKCKCK 306  
300 KANAGKPRAGDGDVITYETFDYVPOULRMFEWMAEDFCRRKKKLEMLEKQGRGKDS 359  
307 KGNKNDQ-----VPTFDYVPOULRMFEWMAEDFCRRKKIIVGIVKTYCRKRYKS 356  
360 -DEYRCSRNIGYDCEQOTISRKGVNMGKCTDCEAFACSGYENWIDNQRKQDPKQ--KYTK 417  
357 GNEPRCSRNIGYDCEQOTISRKGVNMGKCTDCEAFACSGYENWIDNQRKQDPKQ--KYTK 416  
418 EI-----SDGGKRRKRAVGGTITKTEGYSFYEKIKNGYGTVDALFGLLNNKRAKCI 471  
417 EIKIKKNGASGIRQKR--GTTTKYEGYKKFYDKLEKNYVGTGVEFGLLNNKRAKCEV 474  
472 TDGKIKNEKVA--SGGGV-----GGSGSGTSGAGTDEKCGTFYRESEYO 517  
475 KGGGIDTRQVAVSTSGGTAVSAGSSTSGGSAAGSSTSGTINMASOGTFYRESEYO 534  
518 PCPDGCVGHGKGNQWERKTKYKRMKMSKLY--KPIKGMVLLLSLKYVKDMMLLKMW 574  
535 PCPDCGVKRAANGNINRYKDSDECKINILRPKKEEGTKIELKSGEGETE---IKEL 591  
575 KPECLTONSSDGSVYVTGASGENSEKELIYDEKCKKKNHEVQKVNVOGEVEEDD-- 632  
592 EPCFOTQNAKTVG-----VANSGSGTSGSOKLYEDMKY--NDVEKDGDDG--VDDDDLE 643  
633 ---ELKGAAGLICILPNPKKNEVEAKSONNHADIQITFHDFEYVVAHMLKDSIMPTK 689  
644 YRIAVNSSGGLCIL--QKNNGEENGK-----OKTINDFNFVAVAMLDLSIMPTK 693  
690 RLKSCISDGKTKKCRNGKCKDCCEKVAQKQETEMKPIKDHFKTQEGIPRKY--FTTLE 748  
694 KIKGCLNKNAIKCTDKCGDCCEFRWVEQKREEWNTIKHEHFKQTDIPTGLTPDALLE 753  
749 LILKLOFLKEDTBEWNTENSLDAEAEELKHLQKILKLENNLAVVYNACTEQKTLMDKLL 808

754 GYLEKGVLLSTKEAVG---DAKETHTIKQL-----LDETAAGGVVAGKNDTITDKLL 805  
809 NHLELDATKCKDCPLPEEDKSRRGSRADSPDIFIPRPEEKEDDENEDD 3DEYRD----- 863  
806 EDELKANCRCKNCEQKPRGEGGAARNILPGVDTTYVDANEDDLDDED 3DEDDDGSGSD 865  
864 -----DEETAKETTGSAATDTTSLDVCPIYGVKVLTKONESIQDA:SLKYG--GNNS 913  
866 VGGSDVGEVEEETAKEATE-----ETTPLDVCNTVTALEGG--LGEA:ROKYEYGRK 918  
914 RLGMRCVT-----PSEGEPT--TSDDKNGAICVPR 940  
919 FRWAKTISGDNTRGSEBSAGPSRKRHTESDSAVTANGSGEATGKSJDKDGAICVPR 978  
941 RRRRLTY---KKIYDWTATGESPOASGSASSTSGSTPPDSKELLKA FVEASAIETPF 997  
979 RRRKLYLGGFRLTDOT-----AVSSEATQAGTSPQSGPKGDALLTFAFVETAAVETPF 1031  
998 LMRKYEEKKAAVAQE--GAGHGLPRYE---EGSPYDPEDKLKE--GKITDGLRQMFYTL 1051  
1032 LMRKYKOEKKKRNKNEVGAGVLAQTIGTLENSEGEONPOKKLOESGEIDDFLRQMFYTL 1091  
1052 GYRDLIFSGSDTTSVSKDTPESSNDNLKNITVLASGSTDEOREKMKRYKEIKNFKCS 1111  
1092 GDRDLIVAGVD-----DKNGGNITILNAGN-----KEDMERAKTIQDEIDKI 1136  
1112 TERSAENLVSHPO-----TWENNGRYIMHGVNCAITLSDKIDAKGVKKRQKLENDP 1163  
1137 LRKSGSEASGAQKNGSISREKMMWDHAPSIWDMGVNCAITLYNMTDASG--APTOQOEK 1194  
1164 N-LMDEANKKPR-----POYOTNKL--DENSGT--SP-----FTTOTQASDNT 1205  
1195 NALLDDEGKPRKONGNGKDYTGVRLEDENSGTALSPNAPASTASCTTOSSSYSEMT 1254  
1206 PPTTLF-----FYKRPRTYRWFMGESFCREKKRLKQIYKDCG--EG--DVGCSGD 1257  
1255 PPTTLNPKLKDYLKPTTYRYLEBMOQNCFERKKKLLDQIYRECKAYDEL GPRDGKCSG 1314  
1258 GEACDSISTHDIYVSPFNCPCGKGHCSYRKVIERKKIEFKOSNAVQ OQKTATRRN- 1316  
1315 GEDCKONLSKKTDTLPSLPCPGGRHCSFYKKIKIKBOEYEQQAYIEORINYNKMK 1374  
1317 ---GNTFDEKFCITLET--WPAKFLERLKNQPCKTNKE-----YGDIDIDFEKDSKT 1365  
1375 VSESNHDKPEFTNLETKYTDANFLQRLKDGCKNNSSEDDQKVNQYIIFD--DISKD--KT 1432  
1366 FOHTEYCGPRKFTKNC--ONGNCGVGLNGCDGKSIDAKETAKMRS:TTDVMVVSQN 1424  
1433 FGHENYCDPCKSFYVNCNNDHCODNS--NGNCKDKMTIAEKLINGVD:TV--LDMRVLD 1489  
1425 DNTFEGDDLKACQAHANFPGIKRQDVKCGYVCGVDICE-----QTIINERTDKEYI 1478  
1490 SATGFGEDELEBACGASANIFKGIKRWQMGKGVNCGVYNAKPRGNETI:RGEKNDKHI 1549  
1479 QIRALFKRVAENFLEDYKINIKSHICIKGSGSKCINCEKNSKCLEI:WIEKKIAEWEN 1538  
1550 TIRALTYHVNQNFLEDDYKIKIKHISHTCTDOSTCONCKCQVGMIKLQOEWEE 1609  
1539 IKKRFNDYENKQDPYUNKSTLEELIPKIAVY---NDDQNTYKLCVFINSGCTLI:SM 1595  
1610 IKKRFNLQYKMSDEYTPRVSYLETFVLOIGAAANNANDVKKLILSEF:KSCGCSAKTNS 1669  
1596 QNNKENDADIDCMILKLGVAKAKNPGKPSGEBKSDCKEPPPLDPEED-----QNPENL 1649  
1670 ENKKNNDADIDCMILKLGVAKAKNPGKPSGEBKSDCKEPPPLDPEEDLL:JEBQNPNNM- 1728  
1650 EPPKFCP---PTTOPPEEGGCTCGNKEKKKDEKKEESEPAKESGPA:AEBAEPAPAESE 1706  
1729 --PGCFPOQDTEQOQEEENINICTPATYVKKKEEKEKEQOEEEBDEKVP----- 1774  
1707 EETETNPEPPPGTGPAPRPTTPAPRPPDTPPLPRLQADEDFDSTI:QIOT--TLPQVVALAGS 1765

```
Dh 1775 -----PRAPAPASKEKKEKPSQPPRRRTLELLDNPVHTALVSTLAMSVCIGFAA 1826
Oy 1766 IAPLFKKKTKASVGNLFQIOLIPKSDYDIPFLKSSNRYIVYSDRYGKTYIWEQSD 1825
Dh 1827 FTYFLKKKTKASVGNLFQIOLIPASDDIPFLKSSNRYIVASDRYVGYKTYIWEQSS 1886
Oy 1826 -EDKAFMSDDTDVTSSESEYEELDINDIYVPGSPKTYLLEVLPEPGNNVTASGKMP 1884
Dh 1887 GDEKAFMSDDTDVTSSESEYEELDINDIYVPGSPKTYLLEVLPEPGNNVTPT----- 1940
Oy 1885 SDRNDIONDCIPSSKKTNDENMOLKEFISMLONOPNDVPNDYTSNGSSNTNTNTTTS 1944
Dh 1941 SDIPSDIPNSDTP-PPITDDEMNOLKDFISMLONTQNTPEP-----IL 1984
Oy 1945 RHVNDNTNTMSRDNMNEENLLPSIHGNTLSGSEYSYV-----NMVN----- 1989
Dh 1985 HDVNDNTHTPSRHNMDQKPFISIHDRNLFSGEYTYDMFNSGNNINISDSTNSMDS 2044
Oy 1990 --SMNDIPINBNVYSGIDLINDLSGKRPIDYDEVLKKRENELEFGENTTKRTSTON- 2046
Dh 2045 LTSNNHSPYNDKNDLSGIDLINDALSGNH-IDIYDEMLKKRENELEFGQHHPKNTISNR 2103
Oy 2047 -VAKTTSNDPIHNOLEFHAKLDRHDMCEKWKKEKEDILNKLKEBWKENTINSKTYNS 2105
Dh 2104 VVTQTSDDPITNOINLFHKLDRHDMCEKWKKNHERLPLKLEIM--ENETHSGDI--N 2159
Oy 2106 DNKPSHNHVLTDVSIQIDMNPRTKNEITMMDTNOOKSTMDTILDDLEKYNDDPYDYF 2165
Dh 2160 SGITSGHNHVLTDVSIQIDMNPRTKNEITMMDTNOOKSTMDTILDDLEKYNDDPYDYF 2219
Oy 2166 EDDIYHDVYKSSMDIYVHNHNTSNMNVPTKMHIEKNIVNKKKEIFEEXEYPIIS 2223
Dh 2220 KHD-IYDVNDKASEDHINMDHNMKNMNSDVPNTVOIEKNVIYN--QELLQNEYPIS 2275

RESULT 3
096296 PRELIMINARY; PRT; 2197 AA.
AC 096296;
DT 01-May-1999 (Tremblrel, 10, Created)
DT 01-May-1999 (Tremblrel, 10, Last sequence update)
DT 01-May-1999 (Tremblrel, 10, Last annotation update)
DE PREMPL.
GN PFBI055C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shellen S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2, sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Science 282:1126-1132(1998).
DR EMBL: AF001434; AAC71966.1;
SQ SEQUENCE 2197 AA; 249668 MM; D2A92DB54535C143 CRC64;
```

```
Query Match 41.9%; Score 5074; DB 5; Length 2197;
Best Local Similarity 47.1%; Pred. No. 1.3e-268;
Matches 1134; Conservative 277; Mismatches 559; Indels 400; Gaps 87;
```

```
Oy 4 SGGSGGTQDEBDAKVLDEFGQKIVHDEYGEAKNVYSELKSLASTLIG-ETAFVNSMQ 62
Dh 3 SGRKGDPODESVRKMFRIGEDYEVQKSEYVYVSELEKSLAPILGIVSGSTNENCN 62
Oy 63 -TEKTYELLLEANSKRMPCKKDGKGNVDPRSVYKQAGYDNKKAKKCSNGMT-----C 113
Dh 63 LVQDYVYKPYVGNNSRYPCKNLKLGITNEBRPSDTLGGQCTKKIKGNEYSTKSGKDCGAC 122
```

```
Oy 114 APERHLCKNKPMMNSDSSKAKHDLAEVMAAKYEGSEIKTHYPKYSGYSGDPF 173
Dh 123 APERHLCKSH---NLESDITTSMTKLLLEVCMAAKEGSDITHYHQRHEDDSQ 179
Oy 174 MCTMLARSFADIGDILIRGRDYLGNKKKKQNGKETEREKLEBOKLEIFKJHDLKDE- 232
Dh 180 ICTMLARSFADIGDILIRGRDYLGNKKKKQNGKETEREKLEBOKLEIFKJHDLKDE- 234
Oy 223 AOKRYNGDEDENFYKLRDWTANRETYWGAMTSCSKELDINSYRATCNDTGGQBSQTHN 292
Dh 225 AETRY-GSDITNTYQLRDEWYANRATYWEAITC--DVHGSDYPRQTCGDGETTATRVKD 291
Oy 293 KCRDCKDGANAGPKADGDVTTYPTTFEDVYPOYLRFMEEMADFCRKKKKLENEKO 352
Dh 292 KCRCKDEGKKRPG-----SNADOVPTFYVPOYLRFMEEMADFCRKKKKLENEKO 345
Oy 353 CRGKDKDEYR---YCSRNGYDCEBOTSIRKCKVBRGKCTDPCFACSGSYENMIDNORKOF 409
Dh 346 CR-----DYKQNTYCSGNGYDCTRTYKKGKLVIGEHCTGCSYWCRLYESWIDNOKLEF 399
Oy 410 DKQK-KTYKEISDG-----GKKRAVG--GTTKEGYEKSFEYELKMDGCTV 455
Dh 400 LKQKQYETEISNNGSGSGSGVGRNKKRGAGETATNYDGYEKYKELKSEYGV 459
Oy 456 DAFGLLNNEKACKDITD--GGKINPKEVNSGGVYVGGSGTSGASGTNDENKGTFRSE 514
Dh 460 DFLKLNNEVYCKKIDEREKIDFTK-----PADKNSNEGTFTYHSE 502
Oy 515 YCQPCPDGVOHGKGNQWERTKYKMRKSLYPRINGKVLKLSLKVNDMLIKKNW 574
Dh 503 YCKPCPDGVRK--DNQKRDYDGKCTR--GKLYEPASGQPTPIKLSGCKQKEIETKL 560
Oy 575 KEPLTQNSDGSVGYVTTG-----ASGNSKEKEIYDEMCKYKHNVEQY-----NVQ 624
Dh 561 KAFK---DQINGDTNTSNVARGGADSGSKSNSELEYEMKCY--NEQYKVKDKNDEE 615
Oy 625 GEVEEDELKGAGLICLTPMPKKNK EYSEAKSONNHADIGCTPHDFEYVVAHMLKDSI 684
Dh 616 DEDEDEDVKVAKAGSLCTLEN---KHNESRNNSNEPQPKTHDPFYFVIGFNLDSM 672
Oy 685 HWRTKRLKSCISDGTKMCRNGKCKDCFEKVVQKETEKKPIKDHFKTQE----- 736
Dh 673 YWRCK--VNSCINNPKRKRCNECKDDCGCFKEWIGKKKEWENIKKHQTEAFKKNREN 731
Oy 737 -----GIRPGYFTTELKIQFLKEDTEBMTENSLOAELELKHOKIL-----K 784
Dh 732 SGIDMFGSLMDS-ADVVELELELEQLFQDIDKDG-----YGVKELKIKELDEEKK 784
Oy 785 LENENNLAVNAGTEOKTLMKLNHNLNDATKCDPLPEEDSRGRSADPSPDIFP- 843
Dh 785 KQAEVAVVYVADNQKTKITDKLLOHEDDANNL-----KTHKEKEEPTQPK-----PG 835
Oy 844 -----RPEEK-----EDDENEDDEDEVRODEBTAK--ETTEG----- 874
Dh 836 AGPGAPASETGETTLEDEEEDEEEDGADGEVGETVDTDEGETVEYQPVKDTRE 895
Oy 875 -----SATTTTSLOYCPYGVKVLTKDNESLDDACSLKYG--GNNSRL--GWRVYTSGE 925
Dh 896 GEEBEAKKATDTTTSLOVCDTVKNAALT--NNDNLTDACKLKGPGGKRRFPWKKVS--SGE 953
Oy 926 -----PTSSDKNGAICVPPRRRLYI---KKIVDATKTESPOASGSASSTSG 972
Dh 954 KSVATAGSSGATGSGDKGALCVPPRRRLYVGLTLKTSAGTSSESPO--GGSSSNASD 1012
Oy 973 -----STPPDSKKAALLKAFVSALETFFLHRRKEE---KKAVAQE-----G 1013
Dh 1013 VSQNGGDDITTT-----ESLRKWFIEATAETFFLHRRKKEWADQAKKELQNGLLIG 1067
Oy 1014 AGHGLPRYEBGSPREYDEBDKL--REGKIPDGLRQMFPTLGDYRIILSGSNDTISVSKDT 1072
Dh 1068 TGAASL--NLGGDSDSNPOTQLOKSGTIPDLFLRLMFYTLGDIILVGVAD----- 1116
```



0Y	1073	PSSNDINLKNIVLLASSTQOEERKANKKE-1KNFRKS--TERSPNLVSHPOWME	1128
Db	1117	---DKNGNNNTLILASGN-KDEOKMEKIOEKIEOILPTSGNKEKRGONSYNROSILMD	1172
0Y	1129	NGKIKYHGMWCAITLTK--DKIAKGEKKIPKIEJENENIMJDKANKPKRPPOYOTNVKLD	1186
Db	1173	RIAEVWHGHWCALITYDDDNGLKGVYKKPKEIENPEKIMNTTTPKPEDEKYOYOTAKLE	1232
0Y	1187	ENSGTSPRTOTOQASDNTPTTLTHFVKRPYTFRMEEWESGFCREKRRKLOIKOIVDCKV	1246
Db	1233	DESG-----EKRPDSASGCTKLDFELIKRPYFRYLEWGEJENFCCKRTMELGKIKEDC-Y	1285
0Y	1247	ENGVDGKSGDGEACDSISTHDXSTYPSFNRCPGCKHCSYRKMLERKKIEPHKOSNAYG	1306
Db	1286	KNG--GRCSGDGLKCNELVIDEKIEGDLICPTCAHCRFYKMKINTREDFENKOSNAYS	1343
0Y	1307	OOK-----TDATRNNGNTFDEKFECKTLETMPDAAKFLERLKNPCKTNKE--YGGSD	1366
Db	1344	EOKKYEENDSAOKNNG-----VCGTLK--DDAEFLNRLKKNPCKNSEJENKKADE	1395
0Y	1357	IDFEKSKTFQHTHEYCGPCPKFNTCONGNCVSG-LNGNCGDGSIDAKEIANKRSSTT	1415
Db	1396	IDFKKPDPTFDADNCKPCSEFKIKCBNNHNSGGTQCKGCKTTIATLEINIKNTNK	1455
0Y	1416	DYVNRVSDNDTNTEFGDGLKDAQOHANIERGIRKDYWKCGYGVADIC---EQTINERT	1472
Db	1456	EVTMLVSDGSKATFEPFDGISECKDKIGRFGIKRDEMECGKVCYGDICNLKKKNIGES	1515
0Y	1473	DGKEYIOIRALFKRWVNFLEJEDYNNKINDKISHCIKKGECSKINCCENKSKLEJMKIEK	1532
Db	1516	D-KKTIIMKELKMWLEFYLEDYNNKIKHISHCTKNGKSKCIKG-----CVKRWQOK	1568
0Y	1533	IAEMENIKRRENDYERKDDPDY-ANKSILEELIPKIAVANOQDVAIKLAFVENSCKTL	1591
Db	1569	KEEMQOIKERNEDQYKSKTSDEYFVNWKSFLTEWIPRIAYANOQDVAIKLSRFGNSCGSA	1628
0Y	1592	ISNTONKENDADICMLKTLGVAKNCKPGKPSGKSDCKE---PPLPDEED--ONPE	1645
Db	1629	SAISRTNGBEADIDCMIKLEKKIDBCKRKPGENSGQTCNETLTHPLVOYODELEIETE	1688
0Y	1646	ENTL--BEPKRCPTTQPPREKGGEGFCGNKEKKDKKRESEPAKESGPAABEPATA	1703
Db	1689	ENPVGKOPSPCPEVEDKKKEEGEFC-----TPASPA--	1723
0Y	1704	ESEETETNFPEPPGCPAASPPTPAPPTPDPLRLPAQDEFPDSTIIQOTTIPFEVAL	1763
Db	1724	-----PAPAPASPTPPAP-----ADEFPDITIIQOTTIPPIGIALAL	1759
0Y	1764	GSIAFLFK-----KTKASVGNLFQILQIPKSDYDIPFLKSSNR	1803
Db	1760	GSIAFLFKLVYICVVMVYIMCFCIYMKTKKHPY-DLFSYINIPKSDYDIPFLKSPNR	1818
0Y	1804	YTPYVSDKRYKKTYYIYEGSGDDEKAFMSDITDYV-SESEYEBELINDIYVPSPKYK	1862
Db	1819	YTPYVSGYRKRYIYLEGSGDTS--GYTDYHSDIATSSSESEYEBEMDINDIYVPSPKYK	1877
0Y	1863	TLIEVLEPSSNNTTASGAKMTPSDOTRNDONODIPSSKTIPTDENMOLKEFISNMLQOP	1922
Db	1878	TLIEVLEPSSNNTTA-----SDTONDIONODIPSKSDNEMNTLKADDFISNMLQOP	1931
0Y	1923	NDVPNDYSGNSSNTNITTTSHRANVDNNTNTTMSHDNNEENMLPSIHGNTLYSGEYS	1982
Db	1932	KDVPNDYKSGDIPENTQ-----PNTLYPDKPEEKPFITSIDHRLNLNGEYS	1978
0Y	1983	YVNVN-VNSAMDPIINDNNVYSIGDILINDLSGKRPDIYDEVYLKRENELEFGTENTKR	2041
Db	1979	YVNVNNTSMD--KRYVSNVYSIGDILINDLSGKNHDIYDEVYLKRENELEFGTNNHYKH	2037
0Y	2042	TSTOVAKTTSDDIINHOLEFHKMLDRHDMCEKKKNEDILNKLKEMKNENNNSGK	2101
Db	2038	TSIHSVAKNTSDPIINQINLPFTWLDHRDMCEKKENHHERLAKLKEEM--ENETHSGN	2095
0Y	2102	TYNSDNKSRHNVLTNDVYSIOIDMDNPKKNETTNDTNOQDSJYMDTJLIDLEK-YNDPY	2160

[illegible]

QY	459	LGLENKACADITD--GKINFEVNVSGGVYGGSGGTSAGSGINDENKCTFTRSEXC	516
Db	446	LEKLSNEBCTCKVYDEBGGITDFKNV-----SDSTGASGTVNESQSTFYRSKC	496
QY	517	QPCEDCGO--HKSG--NOMERTKYVKMWSKLPR---INGMWLLSLKLVKDMKI	569
Db	497	QPCPYCGKVKVYNNCGSSNEMEKNN-GKCSGKALYEPKPDKGTTITLTKSGKHDD---	552
QY	570	LKKNMKEPCLTQNSDGSVSVYTTGASGNSSEKELJDEMKCYKHNEYOKYNNVOGEVE	629
Db	553	TEELKMFCECDKNDDTINSGSGSGGSGSGSROELJYEMKCYKGEVDVVKGHDEDEE	612
QY	630	DDDELKAGAGCIIIPNPKKKKEVSEAKSONNADIGKTFHDFEYVVAHMLKDSIHMPTK	689
Db	613	DYENWKNAGGCIILAKNKKKEBEGNTESEKREPIDKTNPFYFYVAHMLKDSIHKM-K	671
QY	690	RLAKSICISDGKTMK-RNGCNKKDCCEKVKVOKETEMKPRDKHFKTOEGIPREYFTLE	748
Db	672	KLOKCLTNGRNIKGNKCNKNDCECFRWITQKDEMGKIVOHFKTKITGSGSDNTAE	731
QY	749	LI-----LKLOFLKED---TEENTENSLAEAEELKHLOKTKLENNNLAV	793
Db	732	LIPRDHIVLOYNIQEEFLNGDSEDSSEKSEMSLDAEBAELKHLREIISEDNNQAS	791
QY	794	VNAG--TEOQTLMDKLNHLELNDATKCKDCPLPEEDK-----	828
Db	792	VGGGVTEQKNIMDKLNYEKDEADLCLEIHEDDEEKEKGDGNECIEBGENFRYNPCGE	851
QY	829	-----	828
Db	852	SGMKRYPLVANKVAYOMHHKAKTOLASRAGRSALRGDISLAOFKNGRNGSTLKGQICKIN	911
QY	829	-----SKRSADPS-----PDIFP-----	843
Db	912	ENYSNDSRSGSGGCTCKDODHGGVNRIGTEKSNTEGKKQSYKXNVLPFRHEMCTSN	971
QY	844	-----	843
Db	972	LENDVGSVTKNDKASHSLLDVQLAKTDAEILIRYDDONNIOLTPIDQKQDEAMCR	1031
QY	844	-----	843
Db	1032	AVRSPFADLGIIRGRMDMEDSSSTMDETRLITVKNKIKEHKGIDKPNKPTGDESKRP	1091
QY	844	-----	843
Db	1092	AYKRLADWMEANRHQVYRAMKCATKGIIICPGMVPDYIPORLWMTWEAWYCKAQSQE	1151
QY	844	-----	843
Db	1152	YDKLAKKICADCMKSGDGKCTQGDVDCGCKCAACDKYKEELIKWNEQWRKISDKYMLYLQ	1211
QY	844	-----	843
Db	1212	AKTSTNPGRTVLDDDDPDQYVDFLPIPHKASIAARVLVYRAAGSPETIAAAPTYPY	1271
QY	844	-----	843
Db	1272	STAAGYIHOEIGYGGCOEOTOFCSEKKNHGAITSTYTKENKEYTFOKPPPEYATACDINRS	1331
QY	844	-----RPEEKED-----	850
Db	1332	QTEBPKKKEENVESACKYVEKILEGKNGRITVGECPKESYFDMDCKNNDISHDGACMP	1391
QY	851	-----DEN-----EDDEDEYVD---	863
Db	1392	PRROKCLYYIAHESQTEINIKITDNLKDAFIKTAAEFTPLSMQYKXKNDSEAKILDRGL	1451
QY	864	-----DEETAKE-----TEGS-----	875
Db	1452	IPSOFLRSMWYTFEDGYRDICLMTDISKKONDAKAKADIKGFEKSGDGSKSPSGSLROEMW	1511

[illegible]

```

Db      2536  ETPEDQTPVLKPREBEAVPEPPP-----PRQEKARPRIPQRPPIPTIOLDNPHVLTAL 2589
Oy      1752  QF-TTFPGVALAGSLAFLFLKKTAKASVGNLFQILQIPKSDYDIPPTLKSSNRNRYIPYSD 1810
Db      2590  VMTSLAMSVGIGAFATPTTYTLAKTKTKSSVGNLFQILQIPKSDYDIPPTLKSPRRNRYIPYSG 2649
Oy      1811  RPKGKTYIMEDSDSDKAFMSDPTTDVTSSESEBELDINDIYVPGSKYKTYLLEVYLE 1870
Db      2650  KYRGKRYIYLEDSDSGDGS-GYDHDYSDLTSSSESEYEMDINDIYVPGSKRYKTYLLEVYLE 2708
Oy      1871  P-----SGNNTASGKNTPEDFETENDIONDGIPESSKITDNEGNOLKKEFTSN 1916
Db      2709  PGGNNTTASGNNNTTASGNNNTTASGKNTPEDFETENDIONDGIPESSKITDNEGNOLKKEFTSQ 2768
Oy      1917  MLQNPNDVNDYTSQGSNSTNTNTTSSRHNDVNDNTNTMSRBNNEENLLPSIHGNNLY 1976
Db      2769  YLQSEPTNEPRN-----MLGVNDNNTHTPTSHHNVBEKRFMSINDRNLF 2813
Oy      1977  SGEESYANV-----NKNV-----SKMDIPINDDNNVYSGIDLLINDLSGGKPI 2019
Db      2814  SGESEYNDYDFNFGNNDPINSIDTNSMDLSLTSNHSPYNDKNDLYSGIDLLINDALSGNH-I 2872
Oy      2020  DLYDEVLKKEKMEELNGTGE-NTKRTSQNVAKTNSPIRINOLEFHKWLDRDRHDCERKK 2078
Db      2873  DLYDEVLKKEKMEELFGTKHTHTKHTNTYNAKPARDDPTNQINLFRHKLWDRHRHDCERKK 2932
Oy      2079  NKEEDILNKLKEMWKNENINNSGKTYNSDKPSHHNVLNLTVDYSIQIDMDNPKRKNETNMD 2138
Db      2933  NNHERLPKLAKELM--ENETHSGDI--NSGIPSGNHVNLNLTVDYSIQIDMDNPKRKNETNMD 2988
Oy      2139  TWDKSTMDFTLIDDEKRYNDPYDYFEDDIIYHDVDEKSSMDDIYVDHNNVTSNNMDV 2198
Db      2989  TYPDKSTMDFTLIDDEKRYNEPYDYFEDDIIYHDVDEKSSMDDIYVDHNNVTVNNMDV 3048
Oy      2199  PTKMHTEMIVNNKKEIFEEEPISDIWNI 2228
Db      3049  PTKMHTEMIVNNKKEIFEEEPISDIWNI 3078

RESULT 5
ID 026034 PRELIMINARY; PRT: 2182 AA.
AC 026034:
DT 01-NOV-1996 (TREMBLrel. 01, created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)
DE VARIANT-SPECIFIC SURFACE PROTEIN.
VAR-7.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DD2;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herfelde J.A.,
RA Peterson D.S., Ravetch J.A., Wellem T.E.,
RT "The large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmodium falciparum-
RT infected erythrocytes."
RL Cell 82:89-100(1995).
DR EMBL: L42636; AAA75399.1; -.
SQ SEQUENCE 2182 AA; 248722 MW; 813969A01460BAF8 CRC64;

Query Match 39.8%; Score 4817; DB 5; Length 2182;
Best Local Similarity 45.9%; Pred. No.1.4e-254;
Matches 1096; Conservative 298; Mismatches 607; Indels 388; Gaps 93;
Oy      4  SGGSG---GTDP-----EDAKHYLDGFGQVYHDE--VHGAKNYVELKAGSLSLAS-I 50
Db      11  GGGSSGKKKKKPTSEYIYVSDAKDLDRGKEVYKNGCAKRYIEALNKLNNLTANGR 70

```

QY	51	LGETAFYVKSQTESKYTELTLEIANSKRNPKCKDKGDNDRS-----VKEDAG	99
Db	71	SESETIETCTLTKEVEYERVNGDGCRHPRDKAKNEDVRESFDLGGQYYNRKIKDSOO	130
QY	100	YDNKMKCSNMCTCAPPRRLHLCKNPNMNSDSSAKHDLLAEVMAWKEGESIKTH	159
Db	131	GNKK-----GACAPYRHLHCY--NLESIDTYSTHKLLEVCMAWKEGSINTH	181
QY	160	YPKYSDKPSDPMFCIMLARSFADIDJIRGHDLV--NKKKONGKSTREKLEOKL	217
Db	162	YTOHRTMEDSASOLCTYLARSPADJIDYRGDVLIGYDNKKEO-----RKLEOKL	235
QY	218	KEIFEKIH-DNLKDKEAOKRYNGD-EDPNFYKLREDMWTANRETVGAMTCSKELDNSSY	275
Db	236	KDIEFKIKHDVAKTNGAERTIDDAKGGDFPOLREDMWTSNRETVKALICHAPEANFY	295
QY	276	FRACNDIDGOOPSOETHKCRDMDKGNAAKPRAGODVITYPTVTDYVPOYLRFEEEA	335
Db	296	IKTAON-VGKG--TNQOCH-----IGD-----VPTVTDYVPOYLRFEEEA	335
QY	336	EDFCRKKKKLENEKOCRGDKSDEYRYSRNGYDCEOTISRKGVRVAGKGCOTPFAC	395
Db	336	EDFCRKKKKLENEKOCRGDKSDEYRYSRNGYDCEOTISRKGVRVAGKGCOTPFAC	392
QY	366	GSYENMIDNORKOEDKOK-KYTEIISDGG-----RKRAVGSTTYE--GYEKSEYEL	447
Db	393	RYETWIDNOKKEFLKOKRKETEYISGGSGSKPRTKRAARSSSSDI NGEYSREYKL	452
QY	448	KNDYGYTDAFLGILLNNEKACKDITDGGKINFEVNSGGVGGVGGSGTISGASCTNDEK	507
Db	453	KEVGQADKFLKILNRBGIDOKOPGV--NEKADN-----VDFTEKXV	495
QY	508	GTFYSEYCOCPGCGVOHNGGNOEMERTKYVKMRYSKLYKPLNGKMYLLEKSLKVKXM	567
Db	496	KTFSETEICEPCMPGGLF-KGPRW--KVYGDXTGCSAKTYTDPRNITIDIVLYLPDSQ	552
QY	568	MILKKNKEFCLTONSSDGSVGYVTGTGASGNSKKELYDEMKCYKHIEYOKVNVQGEV	627
Db	553	ONILKKYKNFC-----EKARGGGQIKK-----WQCY-----	580
QY	628	EEDDELKAGAGLCLINPRKKNKVSASAK-----SONNHADIQTFHDFIYYVAHKLDS	683
Db	581	---DEHR-----PSSKNNNVCYGTWDEFTQKOT--VKSYVFWPMDVHMLHDS	626
QY	684	IHMRTKRLKSCI---SOGKTPKCGNCGNKKCDCEPKFVKOKETEMKPIIDHFRTQEGI--	738
Db	627	VEWKTE-LSKCINNNNTGNTCRNNKKCTDCCGQKWEKCKQOEMALIIDHFEGKQDIOQ	685
QY	739	-----PEGYFTTLELILK---LOPLKEDETEENTENSILDAEAEILKHLQILKLE	786
Db	666	QKGLVSEPYG---VLDIVLYKGNLLONIK-DYHDT-----DIKHKIKL--LD	729
QY	787	NENNAVNAAGTEOKTILMDKLNLHELNDATKCK---DCPLPEEDKSRJRSADPSDIFI	842
Db	730	EDAAVAVVLGKDMWTI-DKLQHEKQEAQCKOKOBECEKKAQOESRSASAREDEBT	788
QY	843	PRP-----EKEDEDENEDDEDVYRDETAKEITGSAITDTTTSIDV-----	885
Db	789	QOPADSAAGEVEEEDDDDDYDEDDDDVOVEEGKE--EGVTEVTEVEVEEETVQO	846
QY	886	-----CPVKGVLKKNESLQDACSLEY--GGNNRSRGMVCYTPGSEPTSSDKNAGICY	938
Db	847	BGVKCDIYGR-LFEDDKSLKEACGLKTYGGGKEKPRNMKCYTPSGVSYATSGKDAICY	905
QY	939	PPRRRLIYKIKTIVMATKTESPOASGSSEASTSGSTTPPDK--BALIKAFVESAIET	995
Db	906	PPRRRLIYVGLSQWASR-----GDETTEVSSEATSPASQSESEKIKTAFTESALET	959
QY	996	FFLHMRYKEEKKAVA-OBGAGHGLPRVBEESPEYDEBDIKLE-GKPLPGFLQOMTYGLD	1055
Db	960	FFLHMRYKEEKKRPATDGGAGLSLPEPSPEPDPOTOLOQGVLPFFLQOMTYGLD	1019

QY	1054	YROLFESSNPT--SYKDPSSNDLKIYVLASGTOEOREKMN-----KYREIKFR	1108
QY	1054	YROLFESSNPT--SYKDPSSNDLKIYVLASGTOEOREKMN-----KYREIKFR	1108
Db	1020	YKOLYSSNDTSOTTGKOTPPSSNDLKIYVLASGTOEOREKMKOIAQIKKILNGA	1079
QY	1109	KCSTERSAPNLVSHR--QTMWNGKXYIWMOMCALISKRIAKGYEKK--POKIENTENIM	1166
Db	1080	TSGVPVTKNSVKTPOOTWMENTAKIDIMWAMVICALYKRENDARGISAKIEQKDKLAKYM	1139
QY	1167	DEANKKPRPOYOTYTNKLDENGSTPRTTQTOASSDNT--PTTLTHFYKRTYFRWEE	1224
Db	1140	DEANKNPDIKEYOATYTNKLEDESQ-----AKSNDTIOPTTLKNFVETPTFRWHE	1190
QY	1225	WGESFRCRRKRLKOIYVDCKVENGVDGRCSGDEGCSDISPHDSTYPSFNCPECCGHC	1284
Db	1191	MONSFCEFRAKRLKOIHCHEDEDEGE--KQYSGDGECEBTFSKOYNLODUS--SSCAPC	1248
QY	1285	SSYRMKIERKKIEPHKOSMAYGOOKTD-----ATNNGNTPDEKCEFKTLETPDAA	1335
Db	1249	RLYKWTIEKKKTEYEKQKAYAEQKSNYENBEQKDCQCTQSNMNA--NEBSRLGASPTAA	1306
QY	1336	KFLERLAKGPKTKKEY--GGDD--IDPEKDSKTPQHTIEYCGCPKFKTNCQNGCWSGL	1392
Db	1307	BELOKL--GSCCKNNGYENEDBNKIDFKNPKDKFEAKHSCDPCITGYKQCQNGHC--VESA	1363
QY	1393	NG--NODGKSTDAKEIKAMMSSTTDVYMRSDNDTTFE--GDDLKDAQANHTFGRK	1449
Db	1364	NGECKKNK--ITADIDINKKTDPNGNTEBMVSDSTTFEHLDD-----CKSSGTFKGRK	1417
QY	1450	DVMKGYCVGYDIC--BOTINERTGKKEYIOTRALFKRWENFLEJYDNKINDKISHIK	1507
Db	1418	DEMKANVCYGDICTLEKKIKNOGEBGKITYMKELKKWLEYFLIEDYNRIIRKKIKLCTK	1477
QY	1508	KGECSKTCNGCEKNSKLEWIEKTAEMENIKRPNDOYEKKDOPDYNVASILEELIPK	1567
Db	1478	KEDGCKCIKG-----CIEKMWOEKREKOKINDTLEQKND-----GNLTINEFK	1525
QY	1568	IAYVNDONVYKLC-----VENSNGCGLSTNPNKNENAIIDCMKIKGVAKNCPGKS	1623
Db	1526	FOYRFEKNAIKPCDDGLDOKRTSCGLNSTDNSNGNNNDLVJCLINKLOKIKSECKEHS	1585
QY	1624	GKOSDC-----KEP--PPLPEDEDONPEBNTLEPFCPPTTOPPEEKGCE--TCG	1671
Db	1586	GOTOPCDSNLSCKESTLVEDVDYEQNP--BNKYEQKFCPCDMKEPKKENDDEVGCG	1644
QY	1672	NKEKKDEKKESEEPKAKESGPAEBRAP--TAESEETETN--FPEPGCGPAAPSTPA	1728
Db	1645	GDEEKK--KYVEDSYIEQKEEBAASABEESPPLPEAPKKEENVPKP-----	1690
QY	1729	PTPTPTP-----LBPQDPE--POSTILOTIPRGVALALGSAFLFKKTKTASYN	1781
Db	1691	-----PPKRRKRTKNVLDHRAVYIPALMSSTIMMSIGRAATYTYVLLKTKTSSYGN	1744
QY	1782	LFOLLOIPKSDYDIPTLKSSNRYIYVSDRYKKTUYAMEGSD--EDRYAEFSDDTPTS	1840
Db	1745	LFOLLOIPKSDYDIPTLKSSNRYIYVSDRYKKTUYAMEGSD--EDRYAEFSDDTPTS	1804
QY	1841	SESEYEBELINDIYVCSPKYKTLIEVULPEPSGNNTTASGKTPSDTRNDIOND--GIP	1897
Db	1805	SESEYEBELINDIYVCSPKYKTLIEVULPEPSGNNTTASGKTPSDTRNDIOND--GIP	1853
QY	1898	SS-----KTDENMOUKKETSIMLONOPRNDVNDVOTSSNSTNTNTTTSHRYVNDNTN	1953
Db	1854	NSDPRPTTDEWMOUKKIDTSIMLONTONTPE-----LHDVNDNTN	1898
QY	1954	TTMSRDNMNEELTLPSIHDNLNLSGESEYV-----NMVN-----SMDNPI	1996
Db	1899	PTMSRHNMDOKPFLMSIHDNLNLSGESEYVNDPMSGNPFINISDSTNSMDSLTSNNHSFY	1958
QY	1997	NRDNVNYSGIDLNDLSGKRPIDYDEVYLRKREMLEFCTENTKRTSTON--VAKTINS	2054
Db	1959	NKDNLDYSGIDLNDLSGNH--IDYDEMILKREMLEFCTONHNKNITNSRYVYOTSSD	2017
QY	2055	PIHNOLEPHKHYDRHNDCEKAKKNEIDLNLKLEBMKNENINNSGKTYNSDNKPSHHV	2114

Db	2018	PITNDINLFHKKMLDHRDMCEKKNKNNHELPRYLKELM--ENETSHGDI--NSGITSGNHV	2073
Qy	2115	LNTDVSIGIDMDNPKTKNETINMDTNODKSTMDTILLDLEKYNDEYYDFYEDDITIHVDY	2174
Db	2074	LNTDVSIGIDMDNPKTKNETINMDTNODKSTMDTILLDLEKYNDEYYDFYKHD--IYDYV	2132
Qy	2175	DVEKSSMDIYDHNNVTSNNMDVTFKHHIENIYNKKKEIFEERYPS	2223
Db	2133	NDDKASEDHINMDHKNMNNNSDVPTNVOIEMNVIIN--QELLQNEYPIIS	2180
RESULT 6			
ID	Q9XZB8	PRELIMINARY; PRT: 2710 AA.	
AC	Q9XZB8;		
DT	01-NOV-1999 (TREMBLrel. 12, Created)		
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)		
DT	01-MAY-2000 (TREMBLrel. 13, last annotation update)		
DE	VARIANT-SPECIFIC SURFACE PROTEIN.		
GN	VAR.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
NCBI_Taxid	56833;		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-CS2;		
RA	MEDLINE:99238507; PubMed:10220443;		
RA	Reeder J.C., Cowman A.F., Davenport K.M., Beeson J.G., Thompson J.K.,		
RA	Rogerson S.J., Brown G.V.;		
RT	"The adhesion of Plasmodium falciparum-infected erythrocytes to		
RT	chondroitin sulfate A is mediated by P. falciparum erythrocyte		
RT	membrane protein 1.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 96:5198-5202(1999).		
DR	EMBL: AF134154; AAD29126.1;		
SO	SEQUENCE 2710 AA; 309249 MW; 0AC64F0A3D5BF512 CRC64;		
Query Match 39.7%; Score 4809.5; DB 5; Length 2710;			
Best Local Similarity 38.7%; Pred. No. 4.7e-254;			
Matches 1109; Conservative 308; Mismatches 646; Indels 801; Gaps 80;			
Qy	5	GSSGCTOEDBAKHVLDDEFGQKHNDEHGEAKNYSELGSLASLIDETATVYSMQTE	64
Db	8	GSSGSEDDDKAHVLDISIGEYKVEKKVAANYSQLGTSLNAFENEP---GGOOTA	63
Qy	65	SK-----YTELLENKSKRNPKCKDGNDVDFSVK	95
Db	64	NDPCKLLEYHTNNYKKGHGREHPCKLXEYHTNTYTKHGRHPR---KGE-KRFSDV	119
Qy	96	EQAGYDNKKMKCS--NGMTCAPFRLHLCKNKNFPMNSNDSSKAKHD--LAEVCAAYE	152
Db	120	GSGGEDCDNKKIKSKNNKGACAPYRLRLCVRLNLNISALD--KINNDPLLDVCLALHE	177
Qy	153	GESITHTPKYDSKTPGSDFP--MTMTLARSPADIGDIIIRGDLTLGKKKKQKQKEIER	210
Db	178	GOSTIQDPKYQAOYASSFSPSQICTMLARSPADIGDIIIRGDLTLGKKKEID-----	231
Qy	211	ELLEQCKLEIFKIHNDLCKDEAKQRYNGDEPNFYKLRDMWTNRRFVGMAMTSKEL	270
Db	232	--LENNIKKIKTEKILTPDPAKHHYKDEDPNNFQFLREDMWMNNRDEYWKAITCHAG-	288
Qy	271	DNSSYFRATCNDTGGPSQTHNKKCRDKDKGACNAGKPRAGDGVTIVTFEDYVQYLRLW	330
Db	289	ESDKYFRKTYAGCTG--TGCRCKNDK-----KSNTDPPTLYEDYVQYLRLW	336
Qy	331	FEEMAEDECRKKKKKLEMLEKOCRGKDKSDERYYSNRNGYDCEQITISKKGAVRMGCGTD	390
Db	337	FEEMSEDECRKKKKKIKIKVKNCRG--ERDEKYSYSRNGYDCEKTKRAIGRYRMGNOCTK	394
Qy	391	CFEAGSGYENMTIDNQRQFDRK--KKTYTEISIDGGGRKKRAVAGGTYYKGYEKSEFEKLN	449
Db	395	CLFACNPVEMIDNQRQFDRK--KKTYTEISIDGGGRKKGA--ATIKRGYSKFEKELQS	452

QY	450	DGCTVDAFLGLLNNACKDQITD--GGKLNKEVANSGGVGGSGGTSASGASTDNBKK	507
Db	453	NGRTVAFLLEKLENEBDVCKKYDDEGGGKINFPAEKHD-----DNNNDKTK	497
QY	508	GTFPRSYCOCPDCCVQH-KGDNOMERTYKXKRWKSTKYKPINGKAVLLKSLAKYKX	566
Db	498	GTFPRSYCOPDCCGVKKYVNNKNNKEEHRHDSKCKKNINLTKPKNPEDGTKEITLKSSEG	557
QY	567	MLIKKMMKEFC--LTQNSSDSGVSGVVTYTGASGNGSEKKELYDWMCKYKHNQYKVNQ	624
Db	558	ETELKEKLEBPCNOINRGTTNGG---VANSGSNSBOLEYEMKCYEKXKHLDKYKKE	614
QY	625	GEVEEDD---ELKGAGGCLILPNPRKNEVSAASQNNHADIQKTFHDFEYVVAHMLK	681
Db	615	EDEYDNDYHNVEYETGGGCLILKNQDNREE-NKASQNPDEIQKTFHDFEYVVAHMLK	673
QY	682	DSIHWKRRKLKSLDSQKTKKCNKNGCKDCEFEKVVQ-KTEWKPLKDHFKYQEGEPE	740
Db	674	DSIYWK-KLLEKLENG-TKKCGKNCNNCCGCFEKKVWQKKKEEWDKIKEHFRKQTDIGE	731
QY	741	GYYFTTELELIKLOFLKEDETEENTENSILDAEAEELKHLOKLTLENNLAAVYVNAQTEQ	800
Db	732	WEPDILLEOYLEKGVLLTSIKEGYN-----AKGHRRIEALKRBE--AVAPADCK	782
QY	801	KTLMDKLNLHENDATKC-----KDC-----PLUEBDSKRGSHADSPDIFIR-PE	846
Db	783	KTIMDKLIEHEDKDAKCICIKHNEEKCNOQOQOOPRPTGEGPARSDSPQPPAPASGGG	842
QY	847	EKEDDENDEDEVEDDECTAKETEGSATD-----TTSLDVCPIV	889
Db	843	DEEDDDDDDEBDNDAEENEEVEETEAPAVNGEGEELPPGPAPVPAGPTVADYCKIV	902
QY	890	G-----	890
Db	903	DGILKNDGNSNEISGCPKPKSYPDMDCKKNINDSHSGACMPRRORCLYRDLTOGELRK	962
QY	891	---KVLTK-----	895
Db	963	PEDILTFINCAAKETHFAWHKKYKDNVANAENELSGKILPECFRKQMYTTFGDFRIDFEG	1022
QY	896	-----DNESIQ-----	901
Db	1023	TDISSCRYIKDTSQTIKSLGDOATTEKGDTHIDNKKLQEWMTYIHGPKEIWEGMLCALTN	1082
QY	902	-----	901
Db	1083	GLSESEKKNILQDYSYNKLNNAEKDDCCLEKASKPQLRWYVEMSDFCERKKLEBDY	1142
QY	902	-----	901
Db	1143	EDVCIRAKDIEGCKNNKNSNNSCYKVCKEIENYITGKKTIOYESQBEKFNTERKQKKPEYNS	1202
QY	902	-----	901
Db	1203	YSKDASBYLKDCLDGTCDCKMKKYSIDWYMKNPHTYDMDKLETKCECPPTPKPASK	1262
QY	902	-----	901
Db	1263	PLARSDTYOPAPPKPRAGEGLGRSLPNEPRDTIEEDDDDDDEDEDEGKVEAKADEOV	1322
QY	902	-----DACSILKYGNNSRJGMPQVPPSGEPTSS	930
Db	1323	DRKAKEPPPASQNDVAVCTVEALKKGLDDACTILKSGNNSRLGKCI-PSGKP-GDT	1380
QY	931	DKNGAICVPPRRRLYIKKLVMDATKTESPOASGSEASSTSGSTPPSPKEALKAFVES	990
Db	1381	GSSGACICVPPRRRLYVTLPTKMAEIVAQPOEG-----GNT---BSSDKLREAFIQS	1430
QY	991	AAITFFLIMHRYKEKKAAYAOE-----GAGHGLPVRYESSPREYDEBDKLBGKIPDGFRL	1045
Db	1431	AAVEFFLIMHFMFKKEKREKREKANGELJPGSSSDVG-DONQOTLONGTPIPPDLR	1489

QY	1046	QMF7TLGDYDILFSSGND--TTYSVKDIPRSSNDMLKATIVLJAGSFEJEREKMYKE	1103
Db	1490	LMFVTLGDRDPCIGKTPDIDIVSGDKDMENIKKAIEVSFPSSGTSPLSEKTR--	1546
QY	1104	IKNFKRSTERSAANLVSHQTMWENNGKVIHMGWALTSKDIJAKVYKKPKOTENPE	1163
Db	1547	-----ESMERNKEHILMGICALITYENGADOTKIEKSDVYK	1586
QY	1164	NLMDEANKPK-----PPOYTYVWKIDENSGTSPPRTOTJASNDTP--	1207
Db	1587	KIFGNNTDPPDKLITYTVYDNKGTPEFSKYKYDEVELKEDENNGAASASJLSGD-TPLN	1645
QY	1208	--TLTHFVKRPTFYRMFEWEGESFCREBRKKRLKQIKVDCKVENGVGRCJGDGEACDIS	1285
Db	1646	NPKLDEKFKLPPFRMHLHENGSDFCYKRAQMLKEVBDNCLKRDGRTKRCJGVEGNCIDL	1705
QY	1266	THDVTSTVSEFPCGCGHSGSSYKRWLEKRIEPIHKSOMNAYGOOKTDARNGNFFDEKFC	1325
Db	1706	SOKNITFEKDFPCPSKRCGLYKRWLSRKAERFDKOSNAYGOOKTCY-NGNNHDEKFC	1764
QY	1326	KTLFETWPPAAKFEFLRLNNGPCKT-NKEYG--GDDIDFEKSKTFOHTEGCGPKEFTN	1381
Db	1765	RTLK--DDAAQFLENT--GSGKNDNNEBGRKRDIDLDFKRPQTQOHNDYODPCSGFNT	1820
QY	1382	CONGCGVSLNGNC-DGD-KSIDAKEIAKMSSTTDVVMYVSDNDYNTFBGDLKDAQ	1439
Db	1821	CRNGCKSDDPTNSCKDKDKKIDITAKIDKTDPNENIMLVSDSTSVGEG-DLND-CI	1878
QY	1440	HANLEFKGRKQVMKCGVYGVCDVCEQDNTNINERDGEYQIOIALRHWENTLEBYNKN	1499
Db	1879	KAGIFKGRKQAMKCDNVGCGVYVCKPEKGGKGNOKKILTRALLHRAWETYLEBYNKR	1938
QY	1500	DKISHCIKGEBSGKINCCEKNSKCLEWYKELIAEMENIKRFXNDYENKQDPYANKS	1559
Db	1939	KKLNCIKKMGSTCINGCEKCKCQVGMVITTKREMGKIKORPDQYKKNDSGYPKVT	1998
QY	1560	ILEELIPKIAVVDQDNVYIKLYCFENSGCTLSNTQNNKENDA----IDCMKLKLGK	1614
Db	1999	ILEELIPQIAAATQKGMHKKRLQKLYKSLKCNCTNDEKEGEDIANKDKIDILDLKTKK	2058
QY	1615	AKNCGKRSKGKODCKEAPPLDEDE-QMPEBNTILEPPKFCPPYTPQPEEK--GGEIC-	1670
Db	2059	TESCPTPTSDHTQACVDSYHLEDEBELEEBEENPVTQNICP--QOKETKVEDEDCI	2116
QY	1671	-----GNKEKDKPEKEESEPKEESGPAEEPPAFASE-----EFTN	1711
Db	2117	TDAPDPDYKEEBEKEEBKDEDEEEDEDEDEDEDEDEDEEBSYSDYVDSBETIED	2176
QY	1712	FPEPGTGPAA-PPSTAPAPPTDTPPLRJQAODEPSTLTQTTIPFGVALAGSIAPLF	1770
Db	2177	DEDEYVLTSSHSSSQPKRLPRLDLSPLK-----KMLESTILMMVGIGIAFCFL	2228
QY	1771	LKKKTAKASVGMFOILLQIPKSDVIDPLKSSNRYIPYVSDRKTKGTYIYMEDSDEKXA	1830
Db	2229	LKKKPKSPV-DLLFVLYLVHKDYOTPTPKSSNRYIPYRSQPYKGYIYMEDSDSGHY-	2286
QY	1831	FMSDITVTSSESSEYELDINDIYVPSPKYKTLIEVLYLEPSGNNT-----ASGNT	1883
Db	2287	-YEDTDTTSSSEYELDINDIYVPSPKYKTLIEVLYLEPSGNNTTYESGNTRASGNT	2345
QY	1884	PSDITENDIDONGRIPSSKITDNEWNLKEETISMLNQNOPNDVNDYTSNNSNTNITTT	1943
Db	2346	PSDITONDIDODGPIPSKTLDEWNTLKHDPISMLNQNPNDDVPHOYISGNSSTNITTT	2405
QY	1944	SHHNVDNNTNTTMSRDKNNEENLLPSJHDCONLYSGEYTSYNNVMNYSKADIPINDNV-	2002
Db	2406	SHHNV-----EKKPFIHSDIRLIYLTGEBYNDVMSYNSGNNDL-YNKNKMLY	2451
QY	2003	-----YSGDILNLSGSKRIDYDEVELKRC	2029
Db	2452	SGQNNVYSGIDPISDNNGPIYSDKNDRISDNHHPYSGDILNLTLSGNNKIDYDEVELKRC	2511
QY	2030	ENELGTE-NTKRTSTONVAKTTNSDINHOLEFKRWLDHRDMOEKKNKEDILNLRK	2088

```
||||| : ||| : | | ||| ||||| ||||| : | ||
Db 2512 ENELEGTGHHPRKRTTNHFAPTRDPLHNLNLFHKWLDHRHRCCEKWEHNERLAKLK 2571
QY 2089 EEMNKENINNSGKTYNSDNK-----PSHN-----HYLNTDSTIODMDN 2127
Db 2572 EEMENDT-STGNGNTHPSGNTPTSPISGKOSDIPSDNNHSHIDRYVYLVNTDSTIODHMDN 2630
QY 2128 PKTKNEITNMDTNODKSTMDTILDLLEK-YNDPRYYDEVEDDIYHDVDEKSSMDIYV 2186
Db 2631 PKPIQFTN-----MDTILDLDPKPFNEPRYYDMYDD-IYDVH----- 2669
QY 2187 DHN--NVTSNMNDVPTKHIEMINYNKKETFEETPTSDIWT 2228
Db 2670 DHDPTVDTNAVDPSKVQIEMD-VNTR--LVKEKYPIADWDI 2710

RESULT 7
Q26033 PRELIMINARY; PRT; 2664 AA.
AC Q26033;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE VARIANT-SPECIFIC SURFACE PROTEIN.
GN VAR-2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FC83;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
  Peterson D.S., Ravetch J.A., Wellens T.E.;
  "The large diverse gene family var encodes proteins involved in
  cytoadherence and antigenic variation of Plasmodium falciparum-
  infected erythrocytes."
RL Cell 82:89-100(1995).
DR EMBL; L40609; AAA75398.1;
SQ SEQUENCE 2664 AA; 302412 MW; 6EA2468511703091 CRC64;
```

Query Match 38.98; Score 4711; DB 5; Length 2664;  
Best Local Similarity 39.38; Pred. No. 1.le-248;  
Matches 1116; Conservative 290; Mismatches 639; Indels 796; Gaps 88;

```
QY 2 ATSGSGGTQEDAKHVLDERGKVDHEVGEA-KNYSELKSGLSLASIGETAFYTKS 60
Db 6 AAGGGGDCIDIRSAKHLDISTGKKVYDVHGDALQPSNGKLGTSLA--TFEKAPEGKQ 63
QY 61 MOTE-----SKYTELEANSKRNPCKKGDNDVDRFSVEQAGYDNKKKCSNGM--- 111
Db 64 TSEDPCLDNHETVHTTVJSGYKENPC---KDRPEVRSTYEGAECDKSKIRGNSNKG 119
QY 112 TCAPFRRLHLCNKFPMNNSNDSSKAHDLAEVCAAKYGESIKTHYPYDSKYPGSD 171
Db 120 ACAFFRRLHLCDDHLEHKKHDKIT--RHNLADVCEAAKFEASELEKRYGQYQLNNSDVN 177
QY 172 FPMCTMLARSPADIGDIIRGDLTLGNKKKKQNGKREKLEOKLKEIFKKINDLKKD 231
Db 178 INITELARSPADIGDLYRGDLTGNK-----EKDLEENLRKIFPKKIDINDLDA 229
QY 232 EAQKRYNGDE--DNPFKYLRDMWTANRETVWGMATCSKELDNSSYFATCNDTGGGSPQ 289
Db 230 HVQGHYHDDDGKTKNYKRLNANMEANRQYWKAITGCA--GGTYRQTC---GTG-TW 283
QY 290 THNKCRCDDKGANAGPKRAGDGVTVPTVFDVYPOYLRFWEEMADDFCKKKKKLENL 349
Db 284 TNEKRCRC-----PINDVPTVFDVYPOYLRFWEEMADDFCKKKKKYVDIV 327
QY 350 EKQGRGDKDSERYCSRNGYDCQOTISRKGVKMGKGCIDCFACGSYENWIDNOKQF 409
Db 328 KTCNCRNYSRN---LYCSGNGIDCOETIRVIGHVIGSECSKSYWCRRYKKWIDNOKKEF 384
```

```
QY 410 DKQKYYTEISDGGRRKRAVGTTK---YEYEEKSFYEKLKNGDYVDAPFLGLNNEK 466
Db 385 LKQKKCCENEMLSKSKKQ-----SKTYVYEGYQKEFKILKSENVGGLDFLELNEER 440
QY 467 ACKDIT-DGGKINREYVNSG-GVVGSGSGTSGASGTNDENKCTFYRSEYQCPQDCGV 524
Db 441 EQGFERNLDGKIDFETVDGCGVGAIG-----GASDSNSGNK-TFSSHQYCEBCPGCV 493
QY 525 QHKGNQNERKTYKMKMWSLYKRPINKMVLKSLKYVDMMLKKMKREPCLTQNSS 584
Db 494 E-LIGNEWKERNK-GECKGKRYNIPKGTKHNVIPVLSFGDEHKETLEKIQFC----- 545
QY 585 DGSVSVYTTGASGNSSEKELYDEMKCYKHNEQKVAVQGEVEEDELKGAGLITLP 644
Db 546 -----AESNDSKLTIEQMKCY-----YDKREY-----VCTLE 574
QY 645 NPKKNKEVSEAKSONNHADIQTEHDFEYVYVAHMLDSIHMTKRLKSCISDQTKYCR 704
Db 575 NRNKSSEDPK-----EIQKTFHNFEYFWIRHLNDSIEMRDK-INNCIEKAKEGCK 625
QY 705 NGCNKKCCCFEKWKYQKETEKKPIKDHFKTQEG--TFBGYFTLLELLKLOFLKEDTEE 762
Db 626 NECSTDCGCFQRMWIGKKKEEMGEIKKHKTQDGSIFGNNDPVLBNVLTNDELFDOTE 685
QY 763 NTENS-----LDAEAEELKHLO-----KIL 783
Db 686 AYGNSQKIQIGIKDLAKKTOAADATQOKNTIDLLFEYDSEAEKCKKIOEBQCPKPT 745
QY 784 KLEN----- 787
Db 746 KYRNPYCGNNTYDALAGVAKLOQEAKEQLDRNDSRSALKANASQKYSNOGDPDFKK 805
QY 788 ----- 787
Db 806 NLGCGTQKHSNAIGDSKNPCNNKKEKERNVGEKKWNGEYKMSHTDLYLPRRQFCTSN 865
QY 788 ----- 787
Db 866 LEHLNTKSTGLTSPKALHSILGDLVLAARKGEGDIKTLIENDNRSSICRTMKKSPADIG 925
QY 788 -----ENNL----- 791
Db 926 DIIRGTDLMDINGDATGVQNNLKIDIFSKITEELKQHPDXFNNDKYTNSKHTKLRSW 985
QY 792 ----- 791
Db 986 WEANRDQVWAMTCPTYKNGIQCATPHDYIPQRLRMVWEAEMFCKEQSRILEELLRD 1045
QY 792 -----AVNAGCTEQ 800
Db 1046 CGSGTTGKCNNDKCAKCDKQCEYKTKIQIPWADQWNLISNKYQIYLVQAKIAIANGSTK 1105
QY 801 KTL---MDK----- 806
Db 1106 STTKKDDKKVNDVLFQKLEHANYGTRPPEAHPDRPRRAATSKSDVYETTAGYTHQE 1165
QY 807 -----LNHELNDARKCD--CPLPEDEKSGRSPADSPDI 840
Db 1166 ARTRECLGQNVFCNNNGNNEAFSLTPHEYKHACKENKASPEE--LGRS-DSFDDH 1221
QY 841 FTRPEEKEDDENDEDEVEDRD--ETAKETTES-----ATDTTSTL--DV 885
Db 1222 QTRPEDEVDHSSBEGEDESDEBEKEEVEYVDGADDEKAGAVSQPEASPTTKDQVKKP 1281
QY 886 CPVIGKVLTKDNESLDQACSLKYG-GNNSRLGMRCV-----TP 922
Db 1282 CDIV-KELFSNVDPDLQACSTKYGPKN--YGNRCIPTKTSNDVTGDBGGSRVYASTP 1338
QY 923 SGEPTTSSDKNGALCVPRRRRLYIKIYVMATK-----TESQASGSEA----- 967
Db 1339 --ESGSNSDRKNGATCIPPRRRRLYVGLQEWANKHNTETVSOGAEATGASGAPAPGK 1396
```

QY	968	SSSSGTTTPPDDSKELKALFVSSAILEFFPLMHKKEE----	KAAYOEOAGAGLPRAVE	1022
Db	1397	ESSSGGKETPDSK-----LRTAFIEASAIVEFFMLDRKKEMWAKKAELOÖN-GDLYSSG	1452	
QY	1023	EGSEYEPDEKLEKRGKIPDGLFQÖMYYTGTADYRDILFSSGNDTTSVSKDPFSSSDNLK	1082	
Db	1453	DGDDP-NPÖKKLANGVYIPDFELRMATYTTIGDRIOLLYHÖGN--TJDSGNTGSMNN---N	1506	
QY	1083	IVLLASGSTBOEERKNKKYKEI--KNFRKCSSTERSAPNLVSHPÖTWMENKGYIWMHVC	1140	
Db	1507	IVLEASGNKEDMÖKIDEBILPÖKNGGTPLVPKSS---AQÖPDKMNEHAEJIMKGIC	1563	
QY	1141	ALT-----SKRIKAYBEKKÖKLENEPENLMD-----ANK-----PKPÖYTÖTNY	1183	
Db	1564	ALTYTEKNPPTKSARGE--NKTLEKDEYEFKFPGSADKHGTASTPYGTÖTKTÖYDEKY	1620	
QY	1184	KLDENSGTSPRTÖTOASSNDPTTLTHVHKRPYTFPMEEGESFCFERKKRLKÖYVD	1243	
Db	1621	KLETSÖG-----AKTBSASDTP-LISDYLVRPYTFRIELEÖGÖNFCERKKRLKÖKEE	1674	
QY	1244	CKVENGDVGRCSGDGEACDSISTHDYSTPSEPCÖP--CGKICSSYRMKIRERKKIEFHÖ	1301	
Db	1675	CM--DGSDDKXYSGGEGCÖD---RROFSINVSADLBGRSGGNSCGRYKMKIRKRREVDÖ	1729	
QY	1302	SNAYGÖÖKT---DATRNNG-NTFDEKFCCTLETMPDAKFLERLKNGPCÖTKNEYGGDI	1357	
Db	1730	ANAYSKÖKTYEERGSKAGINDHNKEFCYKLTCTDAAFLRLKNGPCCKKDNEGGNDI	1789	
QY	1358	DEKDSKTPÖHTEYCGCÖPFRKTNÖONGNC--GVSGLNÖCÖDGKSDIAKELAMRSSTY	1415	
Db	1790	NFGTTEETEPRAEKCPCSSFKNCRNGSCRGDDÖTEKÖNGG-TITTGPNFTMGCTE	1848	
QY	1416	DYVAKVSDNÖTNTFEGDDLEDACÖHAINIFKGIARJDKDVMKGYCVGIDEÖFÖNINERTDGK	1475	
Db	1849	DYVAVHVSDDKANEEBGDGEIDECENAGIFTGIRKÖDMKCRKXCGHLTÖKÖEGKALINDÖ	1908	
QY	1476	EYIÖIRALFRWVNFLEDYKNTINDEXISHCIRKGGSGKINCENKSNCKLEKWIETKIAE	1535	
Db	1909	ÖIILYRALLRKWVEYFLEDYKIKKIKKLPCIEINGNSPCINCNCCKNRVEMJIKLKDE	1968	
QY	1536	WENIKKRFNÖYENKÖÖPÖVNVYS-----ILBELIRLAVNÖDÖNVILCYFENSKGTL	1591	
Db	1969	WTKIR--NHYLEKRNKEGDKNVNTSLVTNVEFLVYÖIAANDRKÖTSLDKLTSLGNC	2025	
QY	1592	ISNÖNN--KENDAIDCMLKKLGVRAKNCÖP-----KPSGÖKSÖDKEPPRLDEÖ-Q	1642	
Db	2026	PENSRKRDGENDNADICMLKLETKIHECKTÖHENSVENSDÖPHRNCGGNPP--DÖEDDL	2084	
QY	1643	NPEBNTLEPPRCPTTÖPPEERKGGTCGNKKEKDEKKEESEEPAKEESGPAAEEPAÖT	1702	
Db	2085	LEENPVEÖGFCÖTTPQÖBERP--DDCKGLEKREKKEÖGÖDÖGAEDEGAL-----VPS	2138	
QY	1703	AESEETETNPPRPGGPAAP--PSTPAP-PRPÖTPPRLPÖADEPDSSTILÖTTIPPGV	1759	
Db	2139	G-----PPGSEPEADKGPVPAELIPRÖEPPDLSHPAVYR---SLVSTLMSVY	2184	
QY	1760	ALALGSIAPLEFKKTKRASGNLFÖLÖIPKSYÖDIPTLKSNNRYIYVSÖRGYKGYIY	1819	
Db	2185	GIGRAATTYRYLKKTKSSVGNLFÖLLÖIPKSYÖDIPTLKSNRYIYPTSGYRÖKRYIY	2244	
QY	1820	MEGSDSDKAYAFMSDÖTÖDVYSSSESEKEELDINDIYVPGSPKKTILEVLEP-----	1871	
Db	2245	LEGDSGIDS-GYTDHXSÖDITSSSESEYEEMDINDIYVPGSPKKTILEVLEP-----	2203	
QY	1872	-----SGNNTTASGKTPSDÖTÖNDIÖNGCIPSSKTTDNEWÖLKEETISNLCÖNÖPNDV	1925	
Db	2304	GNNTTASGNNNTTASGKTPSDÖTÖNDIÖNGCIPSSKTTDNEWÖLKEDEFISÖYLÖSEPTYE	2363	
QY	1926	PNDÖTSGNSSÖTNTITTSNHNVDNNTNTMSDNDNEEWELLPTSHDNLIVSGEYSYNV	1985	
Db	2364	PN-----MLGYNVDNNTHTPTSHHNVEEKPFIMSHDNRNLVSGEYINÖDM	2408	
QY	1986	-----NMVN-----SMNDIPIRKDNVNYSGDILNDSLSGKRPIDIDYEVILKR	2028	

Dd	2409	FNSGNPNINISDSTJNSMDLSTSNNSPYNDRKNDLTSGIDLLINDALSGNH-IDIYDEMCKR	2467
Oy	2029	KENELFEGTE-NTKRPTSTONAKVTTTSHSDPIHHOLELFHMLDRHRDMCKRAKNKEDILNKL	2087
Dd	2468	KENELFGKRHKHTKHNTNYAVNAKRAPDDITNOINFHKWLDORHRDMCKRAKNNEHRLPKL	2527
Oy	2088	KEEMKENINNCKTYNSDNKPSSHNVLTUVDVSIOIDMKNPKTKNETNMDTNODKSTMD	2147
Dd	2528	KELW--ENEFTHSGLD--NSGIPSGNHVLNTUVDVSIOIDMKNPKTKNETNMDTNODKSTMD	2583
Oy	2148	TTLDDLEKYNBPYYDFEDEDIIHYDVVDEKSSMDIIVDHNNNTSNNMDVPYTKHIEMN	2207
Dd	2584	TTLDDLEKYNBPYYDFEDEDIIHYDVVDEKSSMDIIVDHNNNTNMMDVPTKHIEMN	2643
Oy	2208	IYNNKKEIFEEEPYPSIDIWNI 2228	
Dd	2644	IYNNKKEIFEEEPYPSIDIWNI 2664	
RESULT	8		
ID	09UOG6	PRELIMINARY:	PRT: 2209 AA.
AC	09UOG6:		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)		
DE	VAR. MALAP2.56.		
CN	Plasmodium falciparum.		
OS	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.		
OC	NCB1_TaxID=5833;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=3D7;		
RA	Devlin K., Bowman S., Churcher C., Harris B., Harris L., Lawson D.,		
RA	Quail M., Barrrell B.;		
RL	Submitted (Feb-1999) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL: AL035475; CAB62897.1; "		
SQ	SEQUENCE 2209 AA; 250779 MW; 7A349FEFEE9F7CB CRC(4;		
Query Match	38.4%; Score 4648.5; DB 5; Length 2209;		
Best Local Similarity	44.0%; Pred. No. 2.2e-245;		
Matches 1054; Conservative 318; Mismatches 664; Indels 357; Gaps			
Oy	6	GSGGTQDEDAKHVLDPEGCQKYHDEV---HGCAKYVSELKGSLSLASLGTAFTATYSKM	61
Dd	4	GSTGTQDDADKAMFPRIQQOYVDELMKRDDADAKYIKELGKLSFASIJLESACTDPCC	63
Oy	62	QTESVTYTELIAN-----SKRNPC-KKDGKGADVRFVKBDAGYDNKI MICSNGM-TCA	114
Dd	64	QTESVTYTELISGSGSVARGRHPCGVNSGKGEDVSFRFKERYSKIDEKILGCSNSEGACA	123
Oy	115	PFRRHLHNKNFPPNNNSDSSKAKHDLLAEYCMAKYEGESIKITHYPKIDSRY-P-GSDPF	173
Dd	124	PYRRSLSTCNKNOKNT-NYSSAKAINLLDLVCIAIANHGGSIKTHLKQYDAEYPGSGSHT	182
Oy	174	MCTYLARSFADIGDIRRDLYLGKKKKKKONGKETEREKLEOKLETIKLHNDL----K	229
Dd	183	TCTALARSFADIGDIRRKDKLYRRDK-----GEKKKLEHNLTIKIKIHSDVTSGS	234
Oy	230	DKEA-QKVYNGGEDNFKLRREDMTANRETVGMATGKSKE--LDNSSYPRATCMTD--GO	285
Dd	235	NKEALOERINGSKE-NYTKLRREDMTANRETVWEAITDODDOKLANASTFRATCSDSGK	293
Oy	286	GP-SOTAHKCRCDKDKAGNACKPRKAGDGDVYLVPTFYFVPOYLWFMEIYMAEDFCRRKKK	344
Dd	294	GFSQANDKCRCRCKKGNFDQ-----VPLTFYFVPOYLWFMEIYMAEDFCRRKKK	343
Oy	345	KLENLEKCRGKDADVEYRCSRNGCYDCEQIISRKGAKRMKGCCTDCFACGSTVENMIDN	404
Dd	344	YVNIVYTCRCRDKDENSESRNGCFDEKTRRAIGKLRYGCGCTDCFACYPYEKMI DN	403



0Y	405	ORKOPDQK-KYTEM1--SDGGKKRAAVGGTTKYEGYKSPYEKIKJNDQYGTVAFLGL	461
Db	404	KKKEKLEKREKYINVINGTSSSSKRTAAAG-SVNGVYKEFYEKLEBGNVNDLJGL	462
0Y	462	LNNEFACDIDN---GGKINEKVENSGGVVGGSGGTSGAGSTNDENKGTFRSEYCOF	518
Db	463	LNNEKACODIDMDKEGKINFKD-----DHQDINNKKNDGTFYRSKYCOF	508
0Y	519	CPDQGVQHK-----GQNMQRKTYKKMWSLYLPRINKMVLKSLKVVQDKMILKK	572
Db	509	CPYCGVKKNNNGSGGSGKNKEEKHESDKTRIKLYPRSGOGGTPIKILKSEGEKEIKE	568
0Y	573	NMKRECU-TONSSDGSVSGVYTTGASGSGNEKKELEDMEXCKYHNNGYQVUNNOGEVEDD	631
Db	569	KIDDCITYTN-----GTDSIDSSILCDPMKYEIDELYEGOEGBEDVD	615
0Y	632	--DEL-KGAGGLCILPNPKK--NKEVSEKSONNHADLOKTFHDFEYVWVNMILKSIH	685
Db	616	RYUDELVEYTGGLCILLKKEKKEDQEKESDAKSONPDELQTFYDFEYVWVNMILKSIY	675
0Y	666	WTKRKLASCIDQKTMKCRNGCNKKCDCEPEKWWOKETEMKPIKDHFTQESIPREGYFT	745
Db	676	WTKKLDKCLONGKKKCGKIKLGNDDCFCFQWREKKKTETYNKHDFVQOKSIPREGYFT	735
0Y	746	TLEILIKLOFLKEDTBEENTENSDAEAEELTKILKILKLENNNLAAVNA-----	797
Db	736	TLEGYLQLEFENBESADOKONSLDAKTOILKILKILDEBKNOBETAGCGRGVAD	795
0Y	798	TEOKTLMKDLNHLBNLATYKCD---CPREBEKSKGRSADSPDLFPIPREPEEKEDENE	854
Db	796	NKKEITMDKLIDYEKTIATEBIEKHKRPDRE-GAGRS-DPSRDSPPSRPAOEVGDSE	853
0Y	855	DDDEBEVANDDE-----TAKETTESANDTTTSLD-	884
Db	854	DEDEDE--DDEEHPDDGKDGDANEERENSHNDQEDKDTLDAVENTEYEGPGRPATPYD	911
0Y	885	-VCPYGVULTKDNHESLQDACSLLYG--GNSRILGRVCYTPSEPTT-SSDKN-CAICVP	939
Db	912	KVCDVSKLFS--GNDRGDACGTRKYDKYGEKRPNNKCI-PSQDKAPSPSDNOSICVP	968
0Y	940	PRRRRLYKIKIVDMATYESPQASGEASSTSGSTTPPDSKEALLKAFVESALETFFLM	999
Db	969	PRRRRLYVGLGLTKWA--SGNTVVSQAOQTPQDGTSPDVK--LRDAFTQSAIETFFLM	1024
0Y	1000	HRXKEEKAVALQDEAGHGLPRVEGSEPEYDPEKLAE-GKIPDGLRQMFYLLGYRDL	1056
Db	1025	HXYKXDEKIEKKQOQKNGLV-ANTSNGKXHQKLEKOSGILPEDFKROMFYLLGYRDL	1083
0Y	1059	FGSGNDTTSVSKDTPSSSNINIKIYLLAGSTEOERKMNKKYKLEKPNKCSSTSPAN	1118
Db	1084	F-----GKDI--SGKKNMDI-----EKKINGLLP-KNSTPSPAKKNPT	1120
0Y	1119	LVSHQPTWENNNGKYIWHGVCALTSKDKIAKVEKKPOKTEPNENLMDANKKPKPOY	1178
Db	1121	-----EWMSONGHEHINNAMICALITYDTNPAAS-DKPTONEKVKELAMDEQNNKPK-NDY	1172
0Y	1179	QYTNKUL-DEMSGSPPTTOTQASSDPTTLTHFYKRPYTRRMVEDEBESCRBRKRL	1237
Db	1173	QYSVTJTGEBEAGSOLQSTOSKDAAREKPRDLSFKRPPEYFRYLEWQONCKERRKL	1232
0Y	1238	KQIKKDCVENGADVGRCSGDEACDSISTDYTSFVSPNCPGCGKSSYRKMERKKE	1297
Db	1233	KDIKECHGDNENITRYSOSGVGEQCKNNLBNPSTFDELDEPYCAKCYCRYYKKMINTKTE	1293
0Y	1298	FHKOSMAVGOQKTDATBRNNGNTFDKEFCCTLFTWPDAAKFLERLKNGPCKTKREYGGDI	1357
Db	1293	YEKQEKIYVOQKDATSDNKNKDYNSNODGKLQYASIESFLEKVO--CK--KDNCEGTI	1348
0Y	1358	DFEKDSTFOFTEYCGCGCPKFKTNCQNGNCGVSGLNGNDDGKSIDAKELIAMBSSTDY	1417
Db	1349	KF-NGGOTFOHTEDECKSCSKFRIKCNDCKCSGNTKVPVCKDGKTPIDAKELIAMINSPOEV	1407

[illegible]



QY	1817	YIMEGDSD - EDKAYAMSTTDVTSSESEYEEELDINDIYVPGSPKYLIEVLEPSSGN	1875
QY	1796	YIMEGDSDSD - EDKAYAMSTTDVTSSESEYEEELDINDIYVPHAPKYLIEVLEPSSGN <td>1855</td>	1855
QY	1876	YIASGKNTSPDRNDONONGIPSSKTTDENMOLKEFISNMLON - QPNDVENDYDTSGNS <td>1934</td>	1934
Db	1856	YIASGKNTSPDRNDONONGIPSSKTTDENMOLKEFISNMLON - QPNDVENDYDTSGNS <td>1914</td>	1914
QY	1935	SYNTNTTTTTRHNVNDNTTWSRDMENELLPSIHONLSEGESEYVNMVMSMDI <td>1994</td>	1994
Db	1915	PLNTQ-----PNTLYFNKPEEKPTTISHDRDLTGTGDISVNHMSNTTMD <td>1961</td>	1961
QY	1995	PLNRDNNVYSGIDLINDSLSGCKPDIYDEVYAKKRENEFGLENTKRTSTOVAKTTSD <td>2054</td>	2054
Db	1962	PYYVSNVYSGIDLINDALNG--DYDYEILKRENELEFGTNHVKQSIHSVAKLTND <td>2019</td>	2019
QY	2055	PLHNOLEPHKMLDRHRDCEKMKKNEKEDLTKLEEKNKENTNNSGKYVNSDKPSHNIV <td>2114</td>	2114
Db	2020	PLHNOLEPHKMLDRHRDCEKMKKNEKEDLTKLEEKNKENTNNSGKYVNSDKPSHNIV <td>2073</td>	2073
QY	2115	LMTDVSIOQDMNPKTKNEITMDTNOQSTMDTILDLLEKENDPYDYFEDDIYHDY <td>2174</td>	2174
Db	2074	LMTDVSIOQDMNPKTKNEITMDTNOQSTMDTILDLLEKENDPYDYFEDDIYHDY <td>2121</td>	2121
QY	2175	DYKSSMDIYDH--NNTVSNMMDVPTKMHLEMTIVNKKKIFEEBYISDTNMI <td>2228</td>	2228
Db	2122	N-----DHDASTVDSNNMMDVPSVQIEMD-VNTR-LVKEKYPIADVMI <td>2163</td>	2163
RESULT	10		
ID	097312	PRELIMINARY; PRT; 2169 AA.	
AC	097312;		
DT	01-MAY-1999 (TREMBLrel. 10, Created)		
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)		
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)		
DE	PEP1120C PROTEIN.		
GN	PEP1120C.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5833;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRATIN-3D7;		
RA	Oliver K., Bowman S., Harris D., Lawson D., Quail M., Barrett B.,		
RL	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AL034559; CAB39061.1; -		
SD	SEQUENCE 2169 AA; 246913 MW; 1D4BAECB7AFAE3BE CRC64;		
Query Match	35.9%; Score 4341.5; DB 5; Length 2169;		
Best Local Similarity	42.9%; Pred. No. 1.3e-228;		
Matches 1025; Conservative 294; Mismatches 656; Indels 413; Gaps	93%;		
QY	13	EDAKYVDFEGQKHYDEVHGEAKNYVELKSGLSIASILGETAFTVKGMSQESKYTELIE	72
Db	19	KDAKELDMIGQTVSHKVRHEDANRGKLYGILT-----QAQFSNKKERVH	63
QY	73	ANSKRNPCKD-----GKGNDVDSVSKEDQAGYDNKCMKCSNGMCAPRR	118
Db	64	IN---NPLCLDYNNTVNTSNVIPCCEKSVSEFSEVSGGECDEKKIKGNSGCAAPRR	120
QY	119	LALCKNFPNNMNSDSKADHDLAEVCGMAAKYEGESIKTHPKYDSYKPGSDFMCTML	178
Db	121	LHVCGRNLGEQIKPH-TIATNHLVLDVCAQAQFBEKSTISGYPRKQYTKXDSGSTITVL	179
QY	179	ARSPFDIGDIIRGDLVYGNKKKKKONGKETEREKLQKLEIFKKIHNL-----KD	230
Db	180	ARSPFDIGDIIRGDLVYGNKKKKKONGKETEREKLQKLEIFKKIHNL-----KD	235
QY	231	KEAQRYNGDDEPNRYKLEBEDMWTNRETVCWAMGCSKELDNSSYFRATC-NDTGGGSGQ	289
Db	236	QALQARYDDDSGNGYQLEBEDMWTNRETVCWAMGCSKELDNSSYFRATC-NDTGGGSGQ	290

[illegible]

DT 01-MAY-1999 (TREMblrel. 10, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
DE PFC0005W, MA3p8.1 PROTEIN.  
OS Plasmodium falciparum.  
CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5633;  
RN [1]  
RP  
RC  
RC STRAIN=3D;  
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,  
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,  
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holtroyd S., Hornsby T.,  
RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,  
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.A.,  
RA Rutter S., Skellern J., Squares S., Squares S., Sultson J.E.,  
RA Whitehead S., Woodward J.R., Newbold C., Barrall B.G.;  
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium  
RT falciparum."  
RL Nature 400:532-538(1999).  
DR EMBL: AL034560: CAB39115.2; -.  
DR INTERPER: IPR002048; .  
DR PROSITE: PS00018; EF\_HAND; UNKNOWN.1.  
SQ SEQUENCE 2209 AA: 250475 MW: 5F644E29B0C0AE38 CRC64;

Query	March	35.66	Score	4302.5	D6	5	Length	2209
Best Local Similarity	41.55	Pred.	No.	1.7e-226				
Matches	1008	Conservative	319	Mismatches	663	Indels	437	Gaps
QY	9	GTODDAKHVLDDEFGOKVHDEVEHEAKNVYSELGSLSLASLIGETAFAIVKSMQT-----	63					
Db	13	GIEDTTHATIHDRIGKIVIEHAKKNAEDYRSQSLKATFEKAPA-----GQOTPCNTC	68					
QY	64	ESKY---TELLIANSKRNPCKKDGKANDVDFSVYKEADGADNKKMKCSITGMTCAPFRRLH	120					
Db	69	ELKTYQMHNVYKGGKKEVPCR---NGTE-KRFSVSVGGECDDSKIKSHIGACAPFTRLN	124					
QY	121	LCNKNFPMNNSDSSKAKHD-LIAEVCMAAKYEEESIKTHPKDYSKYISGDFPMCTMLA	179					
Db	125	LCVRLENI---NNNGKINOTDLADVCLAAIHEDDSIRGHDHX--KE'NDSSQLCTMLA	180					
QY	180	RSPADIGITIRGRLYLCNKKKKONGKETERKELOKLEIETKIHIDNI.KDKEA---OKR	236					
Db	181	RSFADIGIIGIKDLRYGN-----NGK-----DKLEENLKTFIFGIGHGI.KNGKNTGIEE	231					
QY	237	YNGEDDPNFYKLREDMWTANRETYWGAFTCSKELDNSYPATCNDG/GPSOTHNKCR	296					
Db	232	YG--NDPPEFQJREDPMWTANRETYWKAIRCGAPPADADYFIKYVSG---GKTPQKCR	286					
QY	297	DKDKGANAGKPRAGDGYTIVPTVFYDYVPOYLRFMEAEADFCRRKKKILLENLEKOCRGK	356					
Db	287	I-----DPS-VPTFYDYVPOYLRFMEAEADFCRLRKHL:LNNAKNCREK	330					
QY	357	DKSDERYCSRNGYDQCBOTISRKRKVRHGKSCCTCFACSGEYMDINI)KROPKOK--KY	415					
Db	331	HDDGKTLCDLNGPFCOTDANGKMKKTIENDICISCSCHFVMDINI)KKEPKQNKTY	390					
QY	416	TKELISDGGGRKKRVVGGTTKYEG-----YKSFYERLKNIDGYTVAFLGLLNNEKAC	468					
Db	391	DKELIKAVGKN-----GTTTKEISNGPINLNYVGFYSKLQOT--YGSVD:FLKRLINDAIC	445					
QY	469	K---DITDGKINKEVNVSSGGVYVGGSGGTSAGSINDENKKGFFYS.YCOPCPDQGVQ	525					
Db	446	KKHPEVKEKTVNF-----NENLDLFEHPT:YCOACPLCGLE	482					
QY	526	HKGGMQERKTRKVKMKMSKLYKPINGKMYLLLSLYVVDMMILKKNKKEFCFLTONSSD	585					
Db	483	ND-----SPPMK-----KPEKBCRQOQ--RN:PD-----NESN	510					
QY	586	-----GSVGSVYTTGASGSGSEKKEYLDE--WICYKHNEYQKYN/OGVEEDDEL	634					
Db	511	ETKLIVKRCGQOTWIEKJGGJGCGAGKAKNNIOELTWICY--YDKNKNENJ3GGGKD----	564					

OY	635	KGAGGLCLLPNPKKNKEVSEAKSONNHADIQKTFHDFEYYVAHHMLKDSIHMRTRKLSK	694
Db	565	-----YCLIKDKXNNRTOLEIVS-----FNSLFRWMTJEMLKDISIDWR-KEYKNC	608
OY	695	ISDCKTMMCRNGCNKKDDCEFKNVKQETBMRKIXOHFTOGSIPBGY-FTTLELLIKL	753
Db	609	INNGKSTCKNVCKRPDCDQKMGVNRKTKEMOOVKAHY-KEDFGGLPYRTTEIWEJDL	667
OY	754	QFLKEDTEENENSLDAEEAEELKHLOKLITKENENNLAVNAGETOKTLMKLNHELN	813
Db	668	SYPFIKRAHKPE-----KPVOKMELLKKNQEN-----ISRVTKOONSIIRKIFLOQLQ	716
OY	814	DATAC-----KDCLPREDBXSRC-----RSAD-----PSPDIETPREKEDDENEDD	857
Db	717	EANNCLKRKODCKPPQOASABEGAKTGOPRADEEDSDSRPRAGAAGEVDDEDADDANDE	776
OY	858	-----DEBEVRDE-----ETAKEETEGS-----ANDTTS	882
Db	777	ITPRDLNIYUDDLDSKRPDOVEBEKKNDTDETVELAETKEEDTDKRGEKQKPEEYDN	836
OY	883	LDVCPIVEGVLYLT-DNESLDADSLKYG-GNNSRLMRCVTP-----SGEPITSDDNKGAIC	937
Db	837	VKPCGIYVTLFTNGDNTALMBACKQKYQYEKAFPMMKVCPTTRSGSEPRTGS--GSIC	893
OY	938	VPPRRRLRIYIKKIYDMA-----TKTESPOASGEA-----STSGS	973
Db	894	VPPRRRLRIYVPLTRLRLAGGNTAASOGSGEAPVPVTVTO-PASGGNTQVAVSPGAASS	952
OY	974	TTPPSKALLKATVESALATEFPIMHRYTEEKKAQAAGAGLG-PRVEBSGPVEDPBK	1033
Db	953	TSTESSOLLROAFIESANAETFFLMHKRYVDKEKEDKERNRMVVYTSPVN-DLYEK	1011
OY	1034	LKEGPIPDGFLROMFYTLGDYRIDLESNDTYSVKDTPRESSNDMLKNIYLVLASTGE-	1092
Db	1012	LKGGEIPDPFLRQGFYTLADYKDIRESG-----DKDMKNGIND-----ILSGDEL	1057
OY	1093	QERKAMMYKIKIKAFRCCTERSAPMLVSHP-----OTWMENNGKITWHGCVALTISK	1145
Db	1058	HERE----KEIKC---ALEKHFSNVKPTPTSGNDAKTAWKEKNAPIHMEICALTYUK	1108
OY	1146	DKIAKGYEKKROKLENENLMD-----ANKKPKPOUYOTANYKLDENSGT-----	1191
Db	1109	ETSSGGEEK---QIEKSDVIYNKFPGSTAOTKGTIESKTKITDYKLENSGSTGGRPN	1165
OY	1192	-SPRTTOQASSDNTPTTLTHFYKRPTRYFMWBEWSECFREBKRLKOIKUDKVENGD	1250
Db	1166	ESPNSKPPSSSTEUNKPTTLDSFIKRPYFLRYEBWGEFCRBARKLAQIKVDCGDENT	1225
OY	1251	VGRCSGGOGEACDSISTHDYSTVPSCFPGCGHKHCSSYRKWIETERRKIERPHKOSNAY-GOOK	1309
Db	1226	NRSNDGGGFODECEKVTKADVLEDIEDFGSSCATCCSSYRWIERKRIIEEEDENAYAEBOOK	1285
OY	1310	TDARNNGNTEFDKEFCFTLETMPDAKFLERLKNGCJTKNYKEGGD-----IDPEK	1361
Db	1286	KNCVNGNNGKGVNYCCKLE-ENAGFELOKL--GSCK--KOSGEDINGNGENEDXLNRÖ	1339
OY	1362	DSKTFQHTERYCGPCPKERTKQCNGNCVSGLNGCDGKSIDAKEIAKRSSTJDYNAVY	1421
Db	1340	PNVTFERAENCKPCSLIEICKKNGVNGCNPTEKCENGE-TVTALETEKMNLDNGIMDLV	1398
OY	1422	SDNQTJNFEDBDLKDQAQHNIIFKGRKIDVWKGYCVGADVIDEQNINIBRTDGKEYIOIR	1481
Db	1399	SDNCKNEIP-BOLKSSCKDANIIRKIGREDWVGKGRPRDVCVLNFNKHIHDKKNVLIIR	1457
OY	1482	ALFPRWENPLEDNKINDISCHIKKGE--SKCJNGCEKSKCLEMIKIEKIEWENI	1539
Db	1458	TLFRMWLEYFEEDYNRIOKTLAKCIENGKQKBCKCGKGEKMCDCVAKYHVEKEKEMWKI	1517
OY	1540	KKRPENDYENK-DOPDYNVKSLIEBLPKTIAVVDNDVNIKLC---VFENSKGCTLIJSN	1594
Db	1518	RKRLEBYKKNAGSDDYKVSFLSD--POP--YNEYNKAIVKPCDDLMAFERSIHNCNGNS	1573

QY	1535	TQNN--KENDAIDCMKKLKVAKAKNCPGK--PSGEQSQSCKEPPRLPDEED---QNPEEN	1647
Db	1574	SQKDKVDERVDVWVCLLDLKEKKEKCEQKQKNSGNPQCEGSTPPDEEELLEDQEN	1633
QY	1648	TL-----BPPKCPPTTQPPREKGS-----ETCGNKKKEKKKEEKESEPAKE	1690
Db	1634	TVGKEKYGKNNKPAICGVEEQEKKKEBDCDKAVTPDPSDTGNGKEKEDSRSEEEVYSSG	1693
QY	1691	ESGPAAEPAATASSETENFPEPPGCTGPAAPSTPAPPTDTPPPLPAQADEPFDST-	1749
Db	1694	DQSGPAPPPESPQEKAPAPADELPPGERPPK--PAAPTSTPEPAQPL---PSDNTS	1749
QY	1730	-ILQTTTPGVALALGSIATLFLKKTAKSVGLFOILOIPKSDYDIPPLKSSNRYIPYV	1808
Db	1750	DILTTTTPFGIALMTIALFLFKKTKSTI--DLKLVINIPKSDYGMGPPLSPNNRYIPYT	1808
QY	1809	SDRKGGKYYIMEDSDQEDYYAFMSQTTDTV--SSSESEYELDINDIYVPGSKYTLLEY	1867
Db	1809	SGKRGKRYIYLEDSQDTS--GYTDHYSDTSSSESEYEMDINDIYVPGSKYATLLEY	1867
QY	1868	VLEPSGNNTTASGKNTPSDPRNDIONDGIPSS-----KITDNEMNOLK	1910
Db	1868	VLEPSGNNTTASGKNTPSDQNDIONDGIPSDPNTPSDIPKTPSDPTPTDDEDMNTLK	1927
QY	1911	KEPISNMLQN-QPNVDPNDYTSNGSNSTNTNTTSSHVNDNNTNTTMSRDNMEENLLPS	1969
Db	1928	DEFISQYQSQEPNDVNDPKSGNSSTNTNTTTSRHN-----EKKPFIMS	1974
QY	1970	IHDNKLVSGEESYSVNVMMVMSMDIPNRN-----NVYSGIDLINDSL--SGGKRPDIYD	2023
Db	1975	IHDNDLYGEEYNTDMKSNNSGILPSSNRSLSGTVKVPYSGIDLINDSLNSNGQDIDYD	2034
QY	2024	EVLKRENELEFGTEENTKRTSTONVAKTNTSDPIHNOLELEFHKMLDRHRMDCKEKNKEDI	2083
Db	2035	ELKRRKELEFGTHNVQOJSHIVAKLTNDPITHNOLELEFHKMLDRHRMDCKELKNDNR	2094
QY	2084	LNLKEEMKENINNSGKTYNSDNKPSHNVLTNDVSIQIDMDNPKTKNEITNMDTNDOK	2143
Db	2095	LAKLKEEW--ENETHSGNHTSDS---NKTLTVDVSIQIDMDGKPKKEFTN-----	2141
QY	2144	STMDITLDLEKTYNDPIYTYFYEDDIIINDYVEKSMQDIYDVH--NNYTSNNMDVYTK	2201
Db	2142	-MDTILEDLDKYNKP--YYVQOD--IYYDVN-----DHDVSTVDSNAMDIPS	2185
QY	2202	WHIEKNIVNNKKEIFESEYPISDIMN	2228
Db	2186	VQIEMD-VNTR--LVKEKYPIADVMDI	2209
RESULT	12		
P90580			
ID	P90580	PRELIMINARY;	PRT: 2647 AA.
AC	P90580;		
DT	01-MAY-1997 (Tremblrel. 03, Created)		
DT	01-MAY-1997 (Tremblrel. 03, last sequence update)		
DT	01-OCT-2000 (Tremblrel. 15, last annotation update)		
DE	FCG3-VARF11-1 PROTEIN (FRAGMENT).		
GN	FCG3-VARF11-1.		
OS	Plasmodium falciparum.		
OC	Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5833;		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-FCR3:		
RX	MEDLINE=97154495; PubMed=9001213;		
RA	Hernandez-Rivas R., Mattei D., Sterkers Y., Peterson D.S.,		
RA	Wellems T.E., Scherf A.;		
RT	"Expressed var genes are found in Plasmodium falciparum subtelomeric		
RT	regions";		
RL	Mol. Cell. Biol. 17:604-611(1997).		
DR	EMBL: U67959; AAC47438.1; -.		
DT	NON_TRT	2647	2647
SO	SEQUENCE	2647 AA; 303264 MW; 65FD700611348BC7 CRC64;	





Oy	1499	NDKISHCIKKEGSGKJNGCEGNSKSLCEKIEKKIKEMENIKKIRPNDOYENKDDOYAVK	1558
Oy	1499	NDKISHCIKKEGSGKJNGCEGNSKSLCEKIEKKIKEMENIKKIRPNDOYENKDDOYAVK	1558
Oy	1896	RKRLNPCIINGEKALICNG-----CVEQDINHKKREWTJNLS-FNEQY-NQDTERPR	1947
Oy	1559	-SILEELPKIAYVNDODNYIKLVEFENSGKTLISNTQNNK---ENDADIMCLKLGVK	1614
Oy	1948	LRFPEVDILROIATIDRGNGNHGVLKLVKSVKCGNSGNSQKEGNEBDVLCLQKLEKK	2007
Oy	1615	AKNCRGKP--SGEQSGCKRPPRLPDEEO--NPENTLEPRKFCP-----	1656
Oy	2008	AEKCKNDPETSIPQOQCEVSPHNIDEEOPRELEENENTYEHKRICDDVLYKHNHNOER	2067
Oy	1657	----PTTOPREKGEFCGKKEKDEKDEESEPKEESGAABEAPTAEBSEETENF	1712
Oy	2068	LVKNPVLQPTLKRKKKK--KRRKKIKKKKNODRPHNLPCGAFITNTP-----KTKI--	2118
Oy	1713	PEPFGTAPAPSTPABPTPDPRLPRLPOADEPFDSTLIQTTIPRGVALALASIAFLFK	1772
Oy	2119	--PSSG-----KNPMEHPAIVP-----ALVTSILAMSVGIGFAAFYFLK	2158
Oy	1773	KKTASVGNLFQIQLIPKSDYDIPTLKSSKRTIPYVSDRKAKKTYTYEGSDDEDKYAFM	1832
Oy	2159	KKTSKSTIDLLEST--IPKSDYDIPKLSIPKRYIPYSGYGRKRYILEGSGGDS--GYT	2215
Oy	1833	SDTPTDV--SSESEEEELDIDIVVPSPKTYLIEVVLPE-----SGNNTTASGN--PPSD	1886
Oy	2216	DHYSDITSSSESEEEEDIDIVPSSPKTYLIEVVLPESGKLSGNTIPIPSGNNTTASD	2275
Oy	1887	TRANDIQDGIIPSSKITDNEMNOLKKEFISNMLQONDPVNDYTSGNSSTNTNTTTSRH	1946
Oy	2276	TQNDIIPTSDP--PRTDDEWNTLKHDFISNMLQONQKDVNDYTSGNSSTNTNTTTSRD	2334
Oy	1947	NVDNNTWTYTSRDMENLLPSTIHGNIYSGSESYNYNNVNSMNDPIPNDDNVYSGI	2006
Oy	2335	NVDNNTWTYTSRDMENLLPSTIHGNIYSGSESYNYNNVNSMNDPIPNDDNVYSGI	2394
Oy	2007	DLINDLSGKGPDIYDEVLKREKNELEFGTEFNKRTSTONAKTNTSPRIHNOLELPHKM	2066
Oy	2395	DLINDLSGKGPDIYDEVLKREKNELEFGTEFNKRTSTONAKTNTSPRIHNOLELPHKM	2454
Oy	2067	LDRHRDCKEKKKNEKIDTLKLKEEMKE--NINN--SGKTYN--SDNK-----	2108
Oy	2455	LDRHRDCKEKKKNEKIDTLKLKEEMKE--NINN--SGKTYN--SDNK-----	2514
Oy	2109	-----PSHN-----HYLNTVYSIQIDMDNKTKTNEI-----TMDNTN	2140
Oy	2515	IPSGKQSDIPSDNNIHSIDIPYLVNTVYSIQIHHNDKPRINETFTYVDSNPNQVDTTYVDSN	2574
Oy	2141	QDKSTMDPTIIDDLEKYNDPYFYFEDDI--IYHDVDEKSSMDIYVDHNNVTSNNNDY	2198
Oy	2575	PDNSMPTIIDDLEKYNDPYFYFEDDI--IYHDVDEKSSMDIYVDHNNVTSNNNDY	2620
Oy	2199	PTKMHIEKNYNNKKETFEFEYPSIDIANI	2228
Oy	2621	PSKVOIEMDI--NTLEIFEFEYPSIDIANI	2647
RESULT	13		
ID	Q25766	PRELIMINARY; PRT; 2042 AA.	
AC	Q25766		
DT	01-NOV-1996 (Tremblrel. 01, Created)		
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)		
DT	01-JUN-2000 (Tremblrel. 14, Last annotation update)		
DE	3D7VARI (FRAGMENT).		
GN	3D7VARI.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5833;		
AN	[1]		
RP	SEQUENCE FROM N.A.		
FX	MEDLINE=96344414; PubMed=8670911;		
RA	Rubio J.P., Thompson J.K., Cowman A.F.;		

RT "The var genes of Plasmodium falciparum are located in the  
 subtelomeric region of most chromosomes.",  
 RL EMBL J. 15:4069-4077(1996).  
 DR EMBL: U53334; AAB09769.1; -.  
 FT NON\_TER 1 1  
 SO SEQUENCE 2042 AA; 233669 MM; 89726CC777BB5C5A CRC64;

Query Match	33.7%	Score 4073	DB 5	Length 2042
Best Local Similarity	41.4%	Pred. NO. 5.3e-214		
Matches 951	Conservative 293	Mismatches 581	Indels 470	Gaps 83

Qy	139	HDLLAEVQMAKVEGESIKTHYKRYKVSQPGSOPFMCSTIAPSPDIDGIDPRLGYCN	198
Db	12	HNLLIDVCLAQIIEGSGISONHCKHOLSTIPDSFOLCTELASSPADIGDIYNGRLLYGN	71
Qy	199	KKKKQNGKETEREKLQELKIEFKITHDNL--KDEAKQRYNGDEPNFYKLREDMNTA	255
Db	72	NR-----ENDKLEKKLSTYGFYKATKIYDNLVEKKKREATDYK-DDAPDFYQLREDMNL	122
Qy	256	NRETFWGAATCSKELDNSSYFRATCDTGGGSPQTHNKCRCCKDKGANAAGKAGGDVT	315
Db	123	NRODVMATITCDA---HDSRYRMKGAD-GSITESAMROR-----NVA	161
Qy	316	IYVPEFYPOYLRYMPEEAMDEFCRAKKKLEMLEKQSGKSKSBEYRCRNQYDCBOT	375
Db	162	DVPEFYPOYLRYMPEEAMDEFCRAKKKLEMLEKQSGKSKSBEYRCRNQYDCBOT	221
Qy	376	ISRKGAIRMGKGTCTDEFFACSGYENWIDNORQFQOK-KYRKETIS-----DGGG	424
Db	222	ISAHHELVOGEBCCKGCVVCIPGPMIDNQOFEKQKRYNTELNEKHDYETTKELISGR	281
Qy	425	RKRAVSGTTR-TEGYKSTYETKLMDGYDAPFAGLLNNKACK-----DITDGGKINF	479
Db	282	RKRSLSL--TTKNKGYDEEY-KIFDEYPDVDFLLLSKETACEASOPYDPRJISNF	338
Qy	480	KEVNSGGVYGGSGGTSASGTNDENKGFYSEYQCPQPCGVOYHKGNGQMERKTKYK	539
Db	339	K-----NKPNPDIISHTETECQACPMCGM-----	361
Qy	540	KMRMSLYKPIRINGMVLKLSKVAKDMILKKMKMEFCLTGNSSDSGVSQVYTTGA--	596
Db	362	-----TMVLWVNOVKTOQRYNATIKFLKPNLTIONTDIPILPDTTSMNIV	408
Qy	597	-----SGNSEKKELIYDMKCYKINEVOYKAVNGSVEDEDELKGAGGLCI---LNP	646
Db	409	EKYRNPFCNSSDDNNSDQINMWQCH-----YDSKSKSGQDNQNVBSTWQNF	453
Qy	647	KKNRESEAKSQNNHADIOKTFHDFEYVYVAHMLKDISIMPRFKRKSCISDQKT-----	700
Db	454	KKDDKVTI-----SYNAFWKMYSEKLDISIMRAELDKLKNDDKTKGSKKNY	500
Qy	701	MKC---RNGCNKKCDCEFEKVAQOKETEMKPIKIDHFKTQEGIPREGYFTTLELILK--LQ	754
Db	501	KNCINFLKNGNLNKK-----KRRNGRO-KKHFKNKK-----IDRAKICLLKMHKL	543
Qy	755	FLKEDTEENENSLDAEEAF-ELKHIOKLITLLENENMLLVANAAGTEOKTIMKLLNHLN	813
Db	544	VLLNDVFLQ-----DMEKAGDQPOHIAKTELKLNKDEKVNVL-SNMETIPEFLQDEEQ	597
Qy	814	DA-----TKCDDPLPEEDKSRSRSASDPSEDIIPREPEKEDDENEDDEVEDARDE	865
Db	598	DAQCVSNNEKECEFYQKPPDQAPG-GAGPSPSDGTDNLDLDDSDGDEDDVSHVDEE	656
Qy	866	ETAKETTESGA-----TDTTSLDVCPIYGVKVLTKDNESLDQACSL	906
Db	657	EPEPNPEEESSEEBKQVYVADTEAAPVKODTOPKEBEVNPCKYI-BELFKSTKNPEADGL	715
Qy	907	KYGGNNSRLMRVY---TPSGEPT-----SSDKGALCIVPRRRRLVYK	948
Db	716	KYGN---YGMKCVHHTSDKGSEPTKARGHSHVARSADGAPSDKQGALCIPPRRKLTYLH	772
Qy	949	KIYDMATKTESPOASGSEASTSGTTPPDSKALILKAFVESAAITETFFLMHYKKEEKA	1008



[illegible]

Db	1745	KPFI <sup>1</sup> SHDR <sup>1</sup> LYGEE <sup>1</sup> XNN <sup>1</sup> VNM <sup>1</sup> STNS <sup>1</sup> MDIP <sup>1</sup> IS <sup>1</sup> SKND <sup>1</sup> VYSG <sup>1</sup> IDL <sup>1</sup> M <sup>1</sup> - <sup>1</sup> (VCG <sup>1</sup> NHND <sup>1</sup> V <sup>1</sup>	1803
Qy	2023	DEV <sup>1</sup> LKR <sup>1</sup> KNE <sup>1</sup> LEGE <sup>1</sup> - <sup>1</sup> NTKRT <sup>1</sup> STON <sup>1</sup> AK <sup>1</sup> TTNSP <sup>1</sup> IN <sup>1</sup> OLE <sup>1</sup> FHK <sup>1</sup> WLD <sup>1</sup> DR <sup>1</sup> 3DM <sup>1</sup> CEK <sup>1</sup> WRKE <sup>1</sup>	2081
Db	1804	DELL <sup>1</sup> KR <sup>1</sup> KNE <sup>1</sup> LE <sup>1</sup> GT <sup>1</sup> KH <sup>1</sup> HT <sup>1</sup> K <sup>1</sup> NT <sup>1</sup> YNA <sup>1</sup> KAP <sup>1</sup> DR <sup>1</sup> PL <sup>1</sup> NO <sup>1</sup> LT <sup>1</sup> PH <sup>1</sup> W <sup>1</sup> LD <sup>1</sup> RR <sup>1</sup> 3NM <sup>1</sup> CEK <sup>1</sup> WK <sup>1</sup> ND <sup>1</sup>	1863
Qy	2082	DIL <sup>1</sup> NK <sup>1</sup> LKE <sup>1</sup> KEW <sup>1</sup> ----- <sup>1</sup> NKENT <sup>1</sup> INS <sup>1</sup> GK----- <sup>1</sup> TYNSDK <sup>1</sup> RP-----	2109
Db	1864	EPLAK <sup>1</sup> LKE <sup>1</sup> EW <sup>1</sup> END <sup>1</sup> ST <sup>1</sup> SG <sup>1</sup> NKH <sup>1</sup> SD <sup>1</sup> PS <sup>1</sup> GR <sup>1</sup> KLSD <sup>1</sup> PS <sup>1</sup> DN <sup>1</sup> NH <sup>1</sup> SD <sup>1</sup> LH <sup>1</sup> PS <sup>1</sup> DI <sup>1</sup> S3KLSD <sup>1</sup> PS <sup>1</sup> DV <sup>1</sup>	1923
Qy	2110	-- <sup>1</sup> SHNH <sup>1</sup> LV <sup>1</sup> ND <sup>1</sup> VS <sup>1</sup> Q <sup>1</sup> ID <sup>1</sup> MD <sup>1</sup> NPK <sup>1</sup> TNE <sup>1</sup> ----- <sup>1</sup> TNMD <sup>1</sup> QD <sup>1</sup> S <sup>1</sup> FM <sup>1</sup> DT <sup>1</sup> I <sup>1</sup> DD <sup>1</sup> LEK <sup>1</sup>	2155
Db	1924	CCSS <sup>1</sup> NK <sup>1</sup> TL <sup>1</sup> ND <sup>1</sup> VS <sup>1</sup> Q <sup>1</sup> ID <sup>1</sup> MN <sup>1</sup> NPK <sup>1</sup> TNE <sup>1</sup> TY <sup>1</sup> VD <sup>1</sup> SN <sup>1</sup> PN <sup>1</sup> OV <sup>1</sup> DT <sup>1</sup> YV <sup>1</sup> DS <sup>1</sup> PN <sup>1</sup> DS <sup>1</sup> 3MD <sup>1</sup> T <sup>1</sup> IED <sup>1</sup> LEK <sup>1</sup>	1983
Qy	2156	YND <sup>1</sup> PY <sup>1</sup> YVE <sup>1</sup> EDD <sup>1</sup> I <sup>1</sup> YH <sup>1</sup> DV <sup>1</sup> VEK <sup>1</sup> SM <sup>1</sup> DI <sup>1</sup> V <sup>1</sup> VD <sup>1</sup> - <sup>1</sup> NNY <sup>1</sup> SN <sup>1</sup> MD <sup>1</sup> VP <sup>1</sup> TK <sup>1</sup> HI <sup>1</sup> VN <sup>1</sup> NK <sup>1</sup>	2213
Db	1984	YNEP <sup>1</sup> - <sup>1</sup> Y <sup>1</sup> YV <sup>1</sup> LD <sup>1</sup> -- <sup>1</sup> YY <sup>1</sup> DV <sup>1</sup> N----- <sup>1</sup> D <sup>1</sup> H <sup>1</sup> DA <sup>1</sup> ST <sup>1</sup> VS <sup>1</sup> NN <sup>1</sup> MD <sup>1</sup> PS <sup>1</sup> KV <sup>1</sup> 2IEM <sup>1</sup> D-VN <sup>1</sup> K-	2027
Qy	2214	E <sup>1</sup> IFEE <sup>1</sup> EP <sup>1</sup> IS <sup>1</sup> DI <sup>1</sup> MI <sup>1</sup> 2228	
Db	2028	-- <sup>1</sup> LV <sup>1</sup> KE <sup>1</sup> YPIA <sup>1</sup> DL <sup>1</sup> MD <sup>1</sup> 2041	

RESULT	14		
Q26032		PRELIMINARY;	PRT; 3006 AA.
ID	Q26032		
AC	Q26032;		
DT	01-NOV-1996 (TReMBLrel. 01, Created)		
DT	01-NOV-1996 (TReMBLrel. 01, Last sequence update)		
DT	01-OCT-2000 (TReMBLrel. 15, Last annotation update)		
DE	VARAINT-SPECIFIC SURFACE PROTEIN.		
GN	VAR-3.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5833;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FCR3;		
RX	MEDLINE=95330813; PubMed=7606788;		
RA	Su X.Z., Heathcote V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,		
RA	Peterson D.S., Ravetch J.A., Wellens T.E.;		
RT	"The large diverse gene family var encodes proteins involved in		
RT	cytoadherence and antigenic variation of Plasmodium falciparum-		
RT	infected erythrocytes. ";		
RL	Cell 82:49-100(1995)		
DR	EMBL; L40609; AAA75397.1; -.		
SQ	SEQUENCE 3006 AA; 343771 MW; 8FD5FF475F943C74 CRC64;		

Query Match	33.0%	Score 3989	DB 5	Length 3006
Best Local Similarity	32.0%	Pred. No. 3,3e-204		
Matches 1034	Conservative 315	Mismatches 644	Indels 1236	Gaps
QY	8	GGTGGDEAKHHVLDFFGGKVDHEVGGKANNVSLKCSLSLASELTGTAIVTKS---	NOTE 64	
		:  :		
Db	6	GGT-NKTAKEYLDITGEVQDKATIEDLYLRNDLGSLSLQAQPHG-VPI DVANPCDLYTE	63	
QY	65	SKYTELEANSRNPCKADKDGNDVDFRS--VKEQAGYDNKKMKKSGNH-TCAPRRLLH	121	
		:  :		
Db	64	I-HTWVVGKREKENC----RGREKRFSDVLSGGCAKMTIKDSVTNSV GACAPRRLLHV	118	
QY	122	CNKNFPNNNSDSSSKAHDDLLAEVCAAKTEGSIITHYPRKDSYTPGIDFPMCTMLARS	181	
		:  :		
Db	119	CDRLN-ELIKTDENTSHYDLIVDVLTKAEHGESLVKHKHKEIKRK-NINENICTVLARS	175	
		:  :		
QY	182	FADIGDILRGDRDLYLGKKKKKQNGKETEREKELEPOKLEIKFIHNDL-----KDK	231	
		:  :		
Db	176	FADIGDIYRGKDLVLTGVDEBEK-----KKRQLEKLNLLKFFQGIHNDVMTSGRTGKKSA	231	
		:  :		
QY	232	EAQRKINGEDEPNFYKLRDEQWTAANKETVWANGATSCKELDNSYFRATCN-----DTGCG	286	
		:  :		
Db	232	EAQRKYN-DAGKNYKKLRDEMNANANRDQVKAATCD-AADNDNEYFENS:DGILYFNSNQ-	288	

OY	287	PSQTHNKCROCKDKGANA	GRKAGKADGDVITVPTFYEDVVPQYLWFEEMADEDCRKKKKL	346
Db	289	-----C-----	GRNEGK-----VPPNLDVDPQHLWFEEMADEDCRKNITPL	325
OY	347	ENLEKOCRGKDKSDEBYRCS	NGYDCBOTLSRRKGYVMGKCTDCEFPAGSEYENMTDNOR	406
Db	326	KSARKCGGEGNT---	KYCSLNGIDCKRYVEEKDSCSDGNCITACJNITICITADAMLRNOR	382
OY	407	KOFDKOR-KYTKELSDGSG	RRKRAVGGTYYEG-----YKSFYEKLKNDGYG	453
Db	383	NEFEKOKIKYTKETE---	-----KYKSSSDKSNISNKNYYNEFENFGKKEYE	428
OY	484	TYDAFLGLLNNEKACKDIT	DGKTINEFVNSGGVYGGGSGTISGASGNDEKGTFFYS	513
Db	429	TLONFLKLLNKGMYCOE---	KIEEBV-----IDNKDEDVYFHRSS	466
OY	514	EYCOCPCCPGQJOHKG---	-----	529
Db	467	EYCOCPCCPVYQCKGKCTE	DKKNDKRSKIILKILIOSEPEIHYLNSDDXOGDITKTL	526
OY	530	-----	NOMERKTK-----	537
Db	527	EYFCSSTJNVEGRVQKWKY	CNKNSDYVNNCEMNISYSKOSTDANVWLSEVCFHSAKKNLL	586
OY	538	VKKMRMS-KLYKPLP	NGKMYLLLSLAKVYKD-----	566
Db	587	IDTJMEHOLKNCINNTN	VNYTCES-KCICECEYEKWKIRKHEHEWKVXNVFGNNRMSY	645
OY	567	-----	MMLL-----KKNKMEF-----	577
Db	646	IYYNNLSVFPDSEFLFOY	MFALDDEKCKWQFTEDLKKEFEPSTNTPTGKSODATIEPLL	705
OY	578	-----CLTQNSSDG-	SVGSVVTVTGASGSGSEK-----KELYDEMKCYKHNVEYOK	620
Db	706	DHLKDMALTCGDNNNS	SNESCDSYKVKVTKNPGCKNPSASNLYVRKRYRLADEMORYARQOLEK	765
OY	621	-----VNOGE-----	-----	626
Db	766	RGGELNLTGDAATKGT	YRQGGPADFKKNVCSINONHTVNONNBRAYTGOCTGDKDSNGG	825
OY	627	-----	VEEDDELKAGGLCI-----	642
Db	826	VRMKIGTYPWKGROIO	MSAEDITYMPRRKONICTSNLEFYLDQTKDGPLKQSGDKLVNNSFLG	885
OY	643	-----	LPNPKRKEV-----	653
Db	886	DVLSAKMDACKIIELYK	KONKNSNLDPEDNNSACRALYSPADIGDIIRGRDLMDKMS	945
OY	654	EAKS-QNN-----	HADIOK-----TFHDEFY--YVVA--HMKDSIMHRTKR	690
Db	946	DAKRLQTMLEKIEFTW	KIKEELPEDIKKRYDKDGTDKLLREDMEANRHQV-----WRA--	998
OY	691	LKSCISDGTWKCKFNG-	-----	706
Db	999	MKCAIENDKDKKC-NG	PIEDYIPQRLRMFTBMAEMFCKEQSRLYNKLVAADCKSCGKAK	1057
OY	707	-----	CNKKKCCFEKWVYKOKETEMKPIK-----	729
Db	1058	SCQKODSGCTGCKRA	CAONYNKKKIRPMEQMEKIKNNKYAOLYKALDASVNGKESKKTFTAS	1111
OY	730	-----	-----	729
Db	1118	DAKDOQVYHFLAELIR	KSGGKGNKTYVSPPTTPNPLYSAGYIHHELGRTVGCMT	1177
OY	730	-----	-----	729
Db	1238	DIDGCONKYKAGKUYG	CMDCNSOIHHTHAGCAMPRRKORCLYSGLTKIDRKALIEYIPT	1291
OY	730	-----	DHEKTOEG-----IPEGY-----YFT-----	745

Dd	1298	EFKSAIETHFAMDRIKEDNGEAEAEELKNGNLPBEGFKROMYTTGDRIDIFFGNDISTH	1357
Qy	746	-----TLELLIKQFLKEDTEENTENSL-----	DAE 771
Dd	1358	AYISGVSPKVTITIEKENDAKYAANKONNNELLDMDWDHGNDIMWGMICALTHKISDPE	1417
Qy	772	EAEELKHLQKLIKTI-----	785
Dd	1418	KKKEIKKNYSYKKNLESPPKSNKNEVEDFAKKPQFLRMFIMGDEFCQAEKREAKYKVS	1477
Qy	786	-----EENNLAVYNA-----	GFEOKTMD--KLNHE-----811
Dd	1478	DAKDYCGCKNTKNSASVCSACKVYEDYITKKVYETKQKGKFDPAEKITDKESYBSFTKD	1537
Qy	812	-----LMDAKCK-----	DCPLPE--DKSRGRS 833
Dd	1538	ASEYLLKKCLDDTCNCQKQKKNNTIEWMTPNKTYYTNSLEKCEQCPQEPPEPPGEGGAR	1597
Qy	834	ADSPDPIFTRP-----	-----BEKEDEN-----853
Dd	1598	SDSGPRDTPTRPAGSDARSMTVPSPRPAPAGDTVHEVALEVEEEDDEDGDLPEQDIBD	1657
Qy	854	-----EDDDEBEVRD-----	863
Dd	1658	EVAGAEEDLDVGAVARILGRTNSPDEDEDEASEEDDDDDDDADDTTEVTGQGEETAE	1717
Qy	864	-----DETRAKTEGSA-----	TDTSIDVCPYIGKVLTKNESLDOACSLK 907
Dd	1718	DHQTTEYTDQAEKAEADKQGGGETPOKETOPKEVYNPCDVI-KULPTTTEFLKACPKP	1776
Qy	908	Y-GGNNSRLGRCVTPSGEPTTSSDKNGAICVPRRRRLYIKKIYDMATKTESPOASGE	966
Dd	1777	YVNGREKFPMMKCI-----	SSGSDASGSICTPRRRKLYLHKI-----EGVD 1818
Qy	967	AS-STGSGTTPPDSKELLALFAVESALIEFFLMHRYKEEKAVAQEGAGHG--LPKVE	1022
Dd	1819	TTVSDDETTPPIYHDALREAFIQTAAVEFFFLMHRYKKIKKEERQEBELQNTFLLPQAQ	1878
Qy	1023	EGSPEDVE--DKLKGRKIPDGLFOMFYTLGDYDILFSGSNPTTSV--SKOTPSSN	1077
Dd	1879	KVSPEDNPEHPOKKLKJGKCIPEEPFKROMFTYLLDIDYDLGVKADVAQAALSGKNSGSD	1938
Qy	1078	DLKNIVLLASGSTOBEREMKNRYKEIKNFKCSTERSAP-----	NLVSHPQTMWENNGK 1132
Dd	1939	KNIKDI-----	SEKISYIEKSGBOTPPKPGQGTITTKPREMWNKGE 1981
Qy	1133	YIHGMVICALT-SKDKIAKGVKKPKIENPEULMBEAKKRRKPPQYOTYTNKIKDENSEG	1191
Dd	1982	HIMAMICALTHNTDTRQYDQVQKOLFEGKNK-----	TRKNSOYOXQKNV-----2026
Qy	1192	SPRTTQIQAASDNMP--TTLHFHVKPYTFYRFEEMGSEFCRERKRRKLOKIVKCYEN	1248
Dd	2027	-----TISVSSNGSPJONITKLEOPASRPFTPLRLLEMGEFPCRKQHKIERIKRECHHDG	2082
Qy	1249	GDVGRCSGDGBCADSISTHDYSTVPSPNCPGCKHCSYRKMIERKKIIEFHKNASVAGQO	1308
Dd	2083	N--RNCDDGDECKEMQPCADGSEFTLQPCSAKCSKYKKNISRKIDFETQKGALEYQ	2140
Qy	1309	KTDATRNNGNTPFKRECKTLETMPDAKFLERLKNCPCTNKETG--GDDIDPEKSKFTQ	1367
Dd	2141	KKDAEGNNMNY--KEFSKTRJRNNDAAAFNLNKLKNSPCSKNDDSDVOEIKIFDDERKFTFG	2198
Qy	1368	HTEKCGCPKPKTNC--ONGKCGVSLNGNDDGKSTDAEKIKMSSSTDDVNRSDNDT	1426
Dd	2199	HEYCKCPCKSTIKYKCKENNNCDNSKRP--DGRNINISIAEDIERKRSSTODYVMSVSDNT	2257
Qy	1427	NTFEGDDLKDACQAHANFKGIRKDWKCGYVGVDC--EQTININERTDGEK-----	YI 1478
Dd	2258	NGKNFYDLND-CIAGIFKRIREDWKCGBGYVDICTLEKTKNNBERYSAKENDKKNQDI	2316
Qy	1479	QIRALFRWVENLETDYKNTDKISHCIKKGEGSKCINCEKNSKSCLEKWIIEKKTAEEN	1538

Db 2317 LIRLFVKWLESLFEDYKINIKSHCMKNDKSPCINGCQKNCQVEKMEKKSEWCK 2376  
Qy 1539 IKKRFNDQENKDDPD-YVAKSILEELIKRIAVNDQNVILCYFENSKGCTLLISNTON 1597  
Db 2377 VBERYINQYRDKNSEAFVKSFLFPIQIPVYDCKGHSITOLKLLKCKCSEKSEN 2436  
Qy 1598 NKENAIDCMLKKLQKAKNCPKSGEQSDCKEPPPLPDEEDONP-----EENPL 1649  
Db 2437 SNEKDVLCCLKLLEDKAKNCDOASGE-----PCPQTTSEPNDEEDILLFEENFV 2487  
Qy 1650 EPPKFCPPPTOPPEEKGFTCGNKKEKDEKKESEEPAKESGPAEAPAESEFE 1709  
Db 2488 EAPNICPKVEEP-----EPVVEEKCDLAAPAKESSE-----TEENSGG 2527  
Qy 1710 TTFPEPGIGPAPSTAPPTPTPTPLP-----QADEPPD--STIIQTTIFGVVALA 1762  
Db 2528 SNEQNPKRKPEEPPEPPPSSETDPAPPTIQPSQADQPNISIDILISTIFPGIALA 2587  
Qy 1763 LGSIAFLFKKTKKASVGNLFQIOLIPKSDYDIPFLKSNRYIPYSDRKGKTYIYMEG 1822  
Db 2588 LSIYFLFKKTKKSSV-DLRLVNIIPKCEYGMPTLKSNRIIPYASDRKKGKTYIYMEG 2646  
Qy 1823 DSDDEKYPASDPTDVTSESESEYEEELINDIYVPGSPKYKTLIEVVLBSGNNTASGRN 1882  
Db 2647 DSDSGHY--YEDTTPVTSESEYEEELINDIYVPGSPKYKTLIEVVLBSGNNTASGRN 2704  
Qy 1883 TSDPRNDIQNGIRSSKTTDNEBNOAKKEPISNMION-QPNVDPRDYISGNSSTNTIT 1941  
Db 2705 TSDPRNDIQNGIRSSKTTDNEBNOAKKEPISNMION-QPNVDPRDYISGNSSTNTIT 2764  
Qy 1942 TTSRHNVDNNTWTMSRDMENLLPSTIHGNYLGESESYNVNVMNMDPIRDRDN 2001  
Db 2765 TTSRHNVDNNTWTMSRDMENLLPSTIHGNYLGESESYNVNVMNMDPIRDRDN 2811  
Qy 2002 VYSGIDLINDLSGCRPIDIYDEVLRKRENELEGTENTKRTSTONVAKTTNSDPIHNOLE 2061  
Db 2812 VYSGIDLINDALNG--DYDIDYEVLRKRENELEGTENTKRTSTONVAKTTNSDPIHNOLE 2869  
Qy 2062 LPHKULDRHRDCEKKNKEDILNKLKEEMKENINSKTYNSDKPSHNHVLNTDVS1 2121  
Db 2870 LPHKULDRHRDCEKKNKEDILNKLKEEMKENINSKTYNSDKPSHNHVLNTDVS1 2923  
Qy 2122 QIDMONPKTKNETJMDTQDKSTMDTILDLLEKYNDPYRYEEDDIIYHODVEXSSM 2181  
Db 2924 QIDMONPKTKNETJMDTQDKSTMDTILDLLEKYNDPYRYEEDDIIYHODVEXSSM 2965  
Qy 2182 DDIYVDH--NNTYSNMNDVPTKMHLEMTNVNKKKEIFESEYPTSIDWNI 2228  
Db 2966 -----DHDTSTVDNSAMDPYKVOIEMD-VNTK--LVKKEKTYRQVYWDI 3006

RESULT 15  
025733 PRELIMINARY: PRT: 2924 AA.  
AC 025733:  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE PFEMP1 VARIANT 1 OF STRAIN MC.  
GN MCVAR-1 PFEMP1.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OK NCBI\_TaxID=5833;  
RN 111  
RP SEQUENCE FROM N.A.  
RC STRAIN=MALAYAN CAMP (MCJ32 K+C+R+);  
RX MEDLINE=95330812: PubMed=7541722;  
RA Baruch D.I., Pasloske R.L., Singh H.B., Bi X., Ma X.C., Feldman M.,  
Tarasch I.F., Howard B.J.;  
\*Cloning the P. falciparum gene encoding pfEMP1, a malarial variant  
RT erythrocytes\*;  
RL Cell 82:77-87(1995).

DR EMBL: U27338; AAB60251.1; -.  
DR INTERPRO: IPR000345; -.  
DR PROSITE: PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
FT VARIANTE 104 104 S -> T.  
FT VARIANTE 105 105 E -> K.  
FT VARIANTE 139 139 D -> G.  
FT VARIANTE 362 362 K -> E.  
FT VARIANTE 503 503 S -> N.  
FT VARIANTE 581 581 S -> T.  
FT VARIANTE 593 593 M -> V.  
FT VARIANTE 617 617 I -> K.  
FT VARIANTE 631 631 V -> G.  
FT VARIANTE 658 658 D -> E.  
FT VARIANTE 722 722 K -> N.  
SQ SEQUENCE 2924 AA; 335862 MW; 7077553BDAC82B26 CRC4;

Query Match 32.5%; Score 3937; DB 5; Length 2924;  
Best Local Similarity 33.3%; Pred. No. 2, 2e-206;  
Matches 1036; Conservative 293; Mismatches 714; Indels 1066; Gaps 103;

Qy 1 MATSGSGGTODEDAKHVLEDFGQKYHDE-VHGEAKNVYSELKSLSL SILGETAFYVK 59  
Db 1 MCGGNGCGGTCKDKDAKHADLRIGEEYKKEVENDAEKFKALKGNLQZKIGIGELASSPN 60  
Qy 60 SMO-TESKYTELEANSKRNPK-----KDGKMDVDRFSYKEDAGIDNKKMKCSNGM 111  
Db 61 PCKLVEDYVNNRKL--RKRYPCANQRTVRFPSDEYGGQCTFNRLKSDENID----SIG- 112  
Qy 112 TCAPFRRLHLCNKNFPNMSNDSKAKHDLAEVCAAKYEGESITKH;PKYDSKYPGSD 171  
Db 113 ACAPYRRLHLCYNIEMKMK--TSTTKHDLILDYCAAAKYBEDSITKH;TKHELNRPDK 170  
Qy 172 FPMCTMLARFADIGDIIRGRLYLGNKKRKQNGKETEREKLEQKLE;FKRIHDLKDK 231  
Db 171 SOLCTILASPADIGDYGKDLGLGYDDEKDO---ERKLENNLIE;FKKHEHLGTO 226  
Qy 232 EAQKRYNGEDENFYFLREDMTANRETYWAGMTCSKELDNSSYPATY;INDQGGGSOCH 291  
Db 227 DAKDHKKDEE-NYQRLRDMWTANSTYWKAITAG-ESDKYFKTY--CSGE--WTD 280  
Qy 292 NNCRCDDKDGANAAGKPADGDVTVPTVEDYVPOVLRFEEMAEDEPC;IKKKKLENLBK 351  
Db 281 DKCRCKDEEGKN-----ETNEYPTTFDYVPOVLRFEEMAEDEPC;IKKKKLENLBK 331  
Qy 352 QCGKDKSDERYKSRNGYDCDQTLISRKGVKMGKCTDCAFACSGYE;WIDNQRKOPDK 411  
Db 332 NCRG-EKGNR-RYCDLNGVNCETARGAELFVGDDCHRCVSACDRFY;WIDNQRKOPDK 389  
Qy 412 Q-KKYTEISDGGGKRAVAGTTKEG-----YKSFYEKLKNDG;IGYDAFAGLLN 463  
Db 390 ORKKYDEEIN-----KTHGTTITTTGKINNLVGHFYKLRK-YIPRYDLSQKLN 440  
Qy 464 NEKACKDITDGGKINKEVNSGGVYGGSGGSGASGTNDENKGFY;SEYCOQPCPDG 523  
Db 441 DEATICK-----KPRN-----VGNKAST--VDNNEVNTTFS;ITTYCEACPMG 482  
Qy 524 VO-HKGNOMERTKYKKNRWSKLYKPIKNGKVLKLSLKYVKDMIL;KNNKEPCLTON 582  
Db 483 AQKEKNGGMKAKKESCAKKEKRIEIKENSTDI---KILTPKGRSKT;LEKLTFC---- 535  
Qy 583 SSDGSVSVYTTGASGSEKKELEIDEMKCYKHNVEQAKNVVNGVEVEDDEELKGAGGLCI 642  
Db 536 -KDG-----QKIKNDIMKCHDD-----NOTDQTD-----SNDV 566  
Qy 643 LRP-----PKKNVESEAKSONNHADIQTFPHDFEYVVAHMLKDS;IHWBTKRLKSCISDG 698  
Db 567 LGDMGNLTKEDKIMS-----YNAFEMWVHMDMLDS;IKMWDEGR-CTIND 611  
Qy 699 KTKKCRNGCNKCDCEKWKVYKQETEMKPIKDHFKYQESIPRGY-VFTTLELILKQFLK 757  
Db 612 KKGTCIKGKNKCIQKQKWEKKEKTEWGIKIDHFRKQKXIDIPDMWTHDDEPLQTLMLKMDLL 671



Db 2772 ELLDKLKEEMNKDNNNGNINPSGNTPTSPDIPSGKOSDIPSDNNIHSDIPYVLNFDVSIQ 2831  
Qy 2123 IDMDNPKTRKNETITNMDTNOCKSTMDTILDDLEK-YNDPYDYDFEEDDITYHDVDEKSSM 2181  
Db 2832 IHMDNPKPINEFSNMDTYPNSSMDTILEDLDPFENEPIYDVODD-LYDVH----- 2883  
Qy 2182 DDIYVDH--NNVTSNNMDVPTKMHITEMNIVNNKKEIFEEEPISDIWNT 2228  
Db 2884 -----DHDTSTVDTNAMDEPSKVQIEMD-VNTR--LVKEKYPIDLMDI 2924

Search completed: May 1, 2001, 13:08:52  
Job time: 340 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2001, 13:03:32 ; Search time 29.5 Seconds

(without alignments) updates/sec  
2367.160 Million cell

Title: US-09-508-967-1

Perfect score: 12100

Sequence: 1 MATSGSGCGTQDDAKHVLD.....VNKKKEFEFEYPISDIMNI 2228

Scoring table: BIOSIM62

Gap 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

93435

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

# SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	418	3.5	1435	1	EBAL_PLAFC
2	315	2.6	1070	1	PVDR_PLAFC
3	301	2.5	2869	1	RBP1_PLAFC
4	277.5	2.3	1658	1	YMG7_YEAST
5	269	2.2	1630	1	MSPI_PLAFC
6	265.5	2.2	1639	1	MSPI_PLAFC
7	265	2.2	1701	1	MSPI_PLAFC
8	264.5	2.2	1726	1	MSPI_PLAFC
9	262.5	2.2	1726	1	MSPI_PLAFC
10	262.5	2.2	3969	1	HRX_HUMAN
11	258	2.1	1701	1	MSPI_PLAFC
12	251.5	2.1	2476	1	ATRX_MOUSE
13	249.5	2.1	2349	1	TPR_HUMAN
14	249	2.1	1153	1	PVDB_PLAFC
15	245	2.0	2375	1	APRX_HUMAN
16	244.5	2.0	2004	1	MOZ_HUMAN
17	238.5	2.0	1637	1	MRSP_STAAU
18	238.5	2.0	1744	1	TANA_XENLA
19	238	2.0	1073	1	PVDA_XENLA
20	237	2.0	1466	1	SPA2_PLAFC
21	233.5	1.9	1391	1	MSR2_YEAST
22	232.5	1.9	2464	1	MAPB_MOUSE
23	226.5	1.9	1183	1	CNA_STAAU
24	226.5	1.9	1875	1	MLP1_YEAST
25	222.5	1.8	1337	1	DEXT_STRDO
26	221.5	1.8	1682	1	DEXT_PLAFC
27	220.5	1.8	1070	1	PVDB_PLAFC
28	218.5	1.8	3924	1	ANK2_HUMAN
29	217	1.8	2843	1	APC_HUMAN
30	216.5	1.8	1679	1	Y109_YEAST
31	216	1.8	2748	1	Y109_YEAST
32	215.5	1.8	1957	1	YD66_SCHPO
33	213.5	1.8	1018	1	FNBA_STAAU

## ALIGNMENTS

34	213.5	1.8	1233	1	VF16_YEAST	P43597 saccharomyc
35	212	1.8	1370	1	2261_HUMAN	Q14202 homo sapien
36	207.5	1.7	1616	1	P200_MYCGE	Q49429 mycoplasma
37	207	1.7	1500	1	SSP5_STRGN	P16952 streptococc
38	205.5	1.7	1700	1	BAR3_CHITE	Q03376 chironomus
39	205	1.7	1790	1	USO1_YEAST	P25386 saccharomyc
40	204.5	1.7	1139	1	HML1_MYCGE	Q49413 mycoplasma
41	204.5	1.7	1251	1	RBP2_PLAFC	Q00799 plasmodium
42	203.5	1.7	2334	1	WAPA_BACSU	Q07833 bacillus su
43	203.5	1.7	2845	1	APC_MOUSE	Q61315 mus musculu
44	202	1.7	1781	1	AKAC_HUMAN	Q02952 homo sapien
45	199	1.6	1460	1	N159_YEAST	P40477 saccharomyc

RESULT 1  
ID EBAL\_PLAFC STANDARD: PRT: 1435 AA.

AC P19214;

DT 01-NOV-1990 (Rel. 16, Created)

DR 01-FEB-1996 (Rel. 33, Last sequence update)

DE 01-FEB-1996 (Rel. 33, Last annotation update)

OS ERYTHROCYTE-BINDING ANTIGEN ERA-175.

OC Plasmodium falciparum (isolate Camp / Malaysia).

OX NCBI\_TaxID=5835;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=90377299; PubMed=2204835;

RA Sim B.K.L.;

RT "Sequence conservation of a functional domain of erythrocyte binding antigen 175 in Plasmodium falciparum."

RL Mol. Biochem. Parasitol. 41:293-296(1990).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: X52524; CA36756.1; -

DR PIR: S11561; S11561.

KW Antigen.

FT DOMAIN 159 1104 ESSENTIAL FOR BINDING TC

FT VARIANT 1031 1031 E -> V (IN STRAINS FCR-3 AND ITG).

FT SEQUENCE 1435 AA; 167389 MW; 32A4309021BIC3D6 CR; 64;

SQ

Query Match 3.5%; Score 418; DB 1; Length 1435;

Best Local Similarity 17.4%; Pred No. 3.3e-11;

Matches 337; Conservative 259; Mismatches 555; Incls 782; Gaps 85;

QY 392 FRACGSEYEMDINORQKQKRYKEISDGGKRAVGGTGYEGYEKSEYK---L 447

DB 17 YFAKANEYDINKNEKFLDYKKEFNEFD-----KKKGVNQKTDKI:FTFLENKLDIL 70

QY 448 KND-----GYGT---VDAPFLGL---NNEKACADITDGKINKREYNSGGGVGGC--- 492

DB 71 NNSKFNKRRKSYGTDPNDIKNLSINKHNNEDEMFNN-----NQOSLSTSSLRKQNY 123

QY 493 -----SGTSGASGTNDENKGFYREYCOPCDGVQKGNQJERKRYKMRMS 544

DB 124 VIMAVRYSRIISFLDSRIINNGRNTSSNNEVLSNCR-----EKRRGKIDCKKKNDRSNV 179

QY 545 KLYKPIGKRVLLSLKLVK-----DMILTKN-----WKEFCLTON 582

DB 180 CI---PDRRIQCIYVLSITIKTYTKETMKDHFLEASKKESQLLKKNN)KKYNSKFCNDLK 236



OY	583	SSDGSVSVNTTGGAS--GSGNSEKKEJDEMKCYKHNHDOQVYVNGVEEDELKAGGL	640
Db	237	NSFLDYGLHANGNDMDRGISTYAKE-----NKIOEV-----FKGAHG-	273
OY	641	GILPNPKKKEVSEAKSONNHADIOKTFHDFEYVAHMLKDSIH-----	685
Db	274	-----EISHKIKIKNFRKEMWNEFRKMLMEALHSEKNNNNNNKNIPOEOLITQW	323
OY	666	---W-----RTRKLASCISDGKTMKCRN-----GONKKC--DC--FEKWYKQET	723
Db	324	IKEMWGEFLERDRNRSLPKPS-----KCKNNTLYEACEKECIDPCPKYADWAIIRSEF	375
OY	724	EMKPIKHOHFKEOEPIR-----EGYFTTLELILQGLKEDTEEN--TENSJAEAE--EL	776
Db	376	EMHILSKYEYLOK--VKRENAENT-----LIKISEKKDAKVSLLNLNCOAEYSKYDCD	427
OY	835	DPSPDIFIPREEKEDEDEDDEDEYRDEETAKETTESATDTTSLDVCPIGVKYL	894
Db	463	-----DKN-----SVDTNTRY-----	473
OY	895	KDNESLDACSLKGGKNNRSLGRRCYTPSEPSTSDKNGALCVPRRR-----RLY	946
Db	474	-----WEC-----KNPIITLTKO--CVCPRRELCLGINDITY	505
OY	947	IKKIYDWTAKTESPOAGSEASSTGSGTTPPDSEKALLKAFESAI--ETFLMRYKEE	10050
Db	506	DKNLL-----MIKHILALAIYTSRLKRRKYYK	534
OY	1006	KKAAVAGEAGHGLRVEEGSEPYEDPEDKLEKGIIPDGLRQMEFYTLGLDRIILFSGSDT	10655
Db	535	-----DDK-----EVCOKINTKFPADIRDII--GGADY	559
OY	1066	TSYSKOTPSSNDLKNIVILASSSTPEQEBEKNKKKEIKENRKRKSTENSAANLYSHOT	11255
Db	560	W-----NOLSRKIKVGIKINTSKSVHNRK--KNDRKLR-----DE	592
OY	1126	WMENNGKTIYHGWICALTSKDKIAKVEKRPORIEPENMLDBANKKRPAPQOYTWNKL	1185
Db	593	WMKIKYAKDYVN--VISVFPADKTY-----CKEDIEIEN-----	622
OY	1186	DENSGTSEPTTQOASDNTPTTLTHFVKRPYFRFMFEEMGSGFCREKRRKLQKIVDCK	1245
Db	623	-----IPQFFRMFSMGDGYOQDKMKMETLKVCEK	653
OY	1246	VENGVDGRCSGDGACDSISTHDTYSPFSPNCGSGCKHCSSTRKTYERKKITFHKOSNAY	1305
Db	654	EK-----PCEDD-----NCKSKCSNKEWISKKEBYENQAOAY	687
OY	1306	GOOKTDATRRNGNTFDEKFCFCTLETETWDAKFLERLKNRCPKTNKEYGDDIDIDFKDSKT	1365
Db	668	QETO-----KGNMYKMYSEF--KSJKPEYLLKYSSEKSN-----LNFEDERE	729
OY	1366	FOHTEY--GCPCPKFR-----TNCQNGNCGVSGLNGN-----CDGKSIDA	1404
Db	730	ELHSDYKKNKTCMCEVADVPISIIIRNNEQYSGAVREENTETAIHRTFPISEGRKGMQ	789
OY	1405	KEIKAKMSSTTDVVMRYSDNDQTFEGBDILKACOHANIFKGIKRDV--WKSGYCVADI	1462
Db	790	KE-----RDDDSLSKISVSPENSREPTAKDT--SNLTK--LKGVDISMKRAVIGSSP	839
OY	1463	CEQFINERTGKGYIDIRALFRWMENFLBEDYNKINDIISMCIKKGBESKOCINCEKNS	1522
Db	840	NDNINATVEQGN-----ISGV--NS	857
OY	1523	KCLEKWTLEKIAEWENIKRRFNDOYENKDPDYNVAKSILEELIPKIAVYNDQDNIKLCV	1582
Db	858	KPLSDVVRP-----KRELEDO--NSDSESETVVN-----HISPSPIINND	897
OY	1583	FENSKGCTLISNTONNKENDAI-----CMKLKLGAKAKCPKPSGKOSDCKEPP	1635

Dd	898	-DSGSGSATVSSSSSSNGSLSDDDRNGDTPVRTODTANTEVYIRKENADKREDEKG----	953
Qy	1636	LPDEDONPEENTLEPPKFCPTTOPPEEK-----GGETCGNKEKKDKKEESEEP-	1688
Dd	954	-ADEERHSTSE-----SLSPEEKMLTDNEGGSNLHEEYKHEHTSNDNVQSG	1001
Qy	1689	-----KEESPAADEPAPFAESETETNFPEPPTGTGAAPSTPAFPPTPLPRQ	1741
Dd	1002	GIVNNVVEKELDTLENPSSSLDECKAHEBELSEPMLSDDOMSNTPG-----	1048
Qy	1742	ADEPFSTLTOTTFPGVALAGSTAFLFKKTKASVGNLFQILQIPKSDIPTLKSS	1801
Dd	1049	---PLDNITSEETEYI-----SN	1063
Qy	1802	NRYIPVSDRYAKGYTYIWEGSDSDDKAFMSDITDVTSSESEYEELDINDIYPGSPKY	1861
Dd	1064	NBY--KVNR-----EDERTLTRYEDIYALKSMNESDGELY-----	1100
Qy	1862	KTLIEVLPEPSGNNTTAGSKNTPSOTRNDIQMDGPSSKITDNEKNOLKEFTISMLONO	1921
Dd	1101	-----DENDELSTVDNESEDA--EAMKMGD-----TSEMSHNS	1132
Qy	1922	PNDVNDTSGSSNTNTTTSRHNVNONTTMSRDMMEENLLPSTHDGNLXSGERY	1981
Dd	1133	SOHIESDOOKNMKVTGDGTT--HVONEISVPATGEDEX--LRSEKSKIHAEE	1186
Qy	1982	SYNNVMVNSMNDIPINRDNNVYSIGDLINDLSGGKPIDIDYEVILKRKENELFGTEPTR	2041
Dd	1187	RUSHDIKHIN--PEDRSNLTLMKDIRE-----ENERH	1219
Qy	2042	TSTONV-----AKTNSDPFHNOLEFHKWLDHRDCMKKAKEDILNK	2086
Dd	1220	LTNQININSOERDLOKHGFHTWMNLHGCVSERSOINH---SHGNGRODGCNSGVNLN	1276
Qy	2087	LKEENKKNKINNSGCTYNSDNCRPHNHVLANTVDSIQIDMDNKTKNETIMOTNDOKSTM	2146
Dd	1277	RS---NNNFNPISRYNLDYK-----KLDDLIEKRNDSYTTELKRLAEI	1320
Qy	2147	DTILDLE-KYND	2158
Dd	1321	MCEMEISKYC D	1333
RESULT 2			
PVDR_PLAYS			
ID	PVDR_PLAYS	STANDARD:	PRT: 1070 AA.
AC	P22290.		
DT	01-AUG-1991 (Rel. 19, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	DUFFY RECEPTOR PRECURSOR (ERYTHROCYTE BINDING PROTEIN).		
GN	PVDR.		
OS	Plasmodium vivax (strain Salvador I).		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5856;		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=91187056; PubMed=1849231;		
RA	Fang X., Kaslow D.C., Adams J.H., Miller L.H.,		
RT	"Cloning of the Plasmodium vivax Duffy receptor."		
RL	Mol. Biochem. Parasitol. 44:125-132(1991).		
CC	-1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP		
CC	DETERMINANT.		
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.		
CC	-1- SIMILARITY: HIGH, TO P.KNOWLEDST DUFFY RECEPTORS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		

or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch).

CC -----  
 DR EMBL: M61095: AA63423.1; -  
 KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane.  
 FT SIGNAL 1 20  
 FT CHAIN 21 1070  
 FT DOMAIN 21 1007  
 FT TRANSMEM 1008 1025  
 FT DOMAIN 1026 1070  
 FT CARBOHYD 183 183  
 FT CARBOHYD 255 255  
 FT CARBOHYD 351 351  
 FT CARBOHYD 420 420  
 FT CARBOHYD 715 715  
 FT CARBOHYD 787 787  
 FT CARBOHYD 825 825  
 FT CARBOHYD 903 903  
 FT CARBOHYD 938 938  
 SQ SEQUENCE 1070 AA; 119683 MW; CB051DF13E294603 CRC64;

Query Match 2.6%; Score 315; DB 1; Length 1070;  
 Best Local Similarity 19.6%; Pred. No. 7.5e-07;  
 Matches 196; Conservative 124; Mismatches 344; Indels 334; Gaps 40;

OY 5 GSGSGTQEDDAKHVDFEGQKVDHVEHGRK--NYVELKGSLSLASILGETAFYVKSQ 62  
 DB 112 GGSYMEKSDG-----GDKTEERKDEHKTDSTKDNKGANNLMVDLETSSNGQFAG 164  
 OY 63 TESKYTELI---EANSKRNPCKRDKGNDV-----RFSVKQAGYDNKK 104  
 DB 165 TLIDNLEFVYTGHEGSRKN--SSNGGNPYDIDHKTTISSAIINHAFLQNTVMKNCYTKRR 223  
 OY 105 MK---CSNGM--TCAPFRRLHLCKNRPNMNSNDSSKAKHD-----LAEVCMAKYE 152  
 DB 224 RERDMDCWTKKQVCIPIRRYOLCKMELTIVLVNNTDTNFHRDITFRKLYLKRILYDAVE 283  
 OY 153 GE-SIKTHYPKYDSKYPSPDFPMCMILARSPADIGDIIRGRDLY-LGNKKKQNKETPR 210  
 DB 284 GULLIKLNNYRIN-----KDF--CQDIRMSLDGFDIIMGTMEGYSK----- 326  
 OY 211 EKLQRLKEIFEKIHNDLKKEAQRKRYNDEDPNFYKLEDMWTANRETVMGAMTCSKEL 270  
 DB 327 -VVENNLISIFG-----TDEKAQR-----RKQMMNESKQOIWAMMYSVK 367  
 OY 271 DNSSYFRATCDNTGQGPSPOTHNKCRDKDGANAGKPRAGDGVITVPTFYDVPQYLEW 330  
 DB 368 RLKGNFIWICK-----LNVAVNIEPQIYRW 392  
 OY 331 FEEMADECCRKKKKKLENEKCRKDKSDERYCSRNKYDCEQITISRKGYRMKGCTD 390  
 DB 393 IREWRDRVYSELPTVEVQKKEKCDGKINITYDKVCK-----VPP 431  
 OY 391 CFFAGSYENWIDNQRFQKQKTYKTEISDGGRRKRAVGSTTKY-----EGEYS 442  
 DB 432 CONAKKSTYDQWITTRKKKNQWDLVS--NKPLSYNAEKVQAGIYPTPDILKQELDERNEVA 489  
 OY 443 FYEKL-KNDGYGTDAFLGL-----NNEKACKDITDGGKINFEVN----- 483  
 DB 490 FENEINKRDG-----AYIELCYCIVEAKKNTQEVYTVNDNAKSGATSNPISQVDS 544  
 OY 484 -----SGGGVVGSGSGTSASGTNDENKGTFFRSEYQCPDCGVQKRGKQWR 534  
 DB 545 KAEKVPDGTGHVNSGDDSTTGRAVTD-----GONGNTPA 583  
 OY 535 KTVVKMRMSKLYKPPINGKAVLLLSLKLVKD-----MMILKKMKKEFLTQNS----- 583  
 DB 584 ESDVQR-----SDIAESVSAKNVDPQKSVSKRSDPTASVTGIAAGKRNGLNASRPSSESTV 640  
 OY 584 -----SDGSVSVVTGASGNGSEKKELYDEWKYKHNVEVQVYVNGVEEEDDELKAG 638  
 DB 641 EANSQDDPTVNSASIPVSGEN-----PLVTPYNGLRHSK-----DNSDSPG 682

OY 639 GCLILPNPKKNEVSEAKSQNNHADIQKTFHDFEYVVAHMLKDSIHMTKRLKSCISDG 698  
 DB 683 PAESNANPDSNSKGETGKQDND-----NAKATKDS-----SNSSPG 719  
 OY 699 KTKMCRNGCNKKKQDFEKVWQKETEWPRIKDHFKT-----OE 736  
 DB 720 TS-----SATGDTTDAVDREINKGPE-----DRDKTVGSKDGGEDNSJNKDAATVYGSD 770  
 OY 737 GIPEGYFTTELLILKQLEKEDTEENTENSIDAEAEELKILQIKLEEN----- 787  
 DB 771 RIRENSAGGSTN-----DRSKNDTERKNGASTPDSKQEDATALSITESLESTSGDRITN 825  
 OY 788 --ENNLAVYNACTEOKTLMDKLLNHELNDATCKCQCPLEEDKSGRSADPSDFIFPR 845  
 DB 826 DTTNSLENKNGEK-----DLQKDF-----KSDTPNEE-----P 857  
 OY 846 EEKEDDENEDDEDEYRDE-ETAKETTESGATDTTTS 882  
 DB 858 NSDQTTDAEGHGRDSTKNDKARERKHMNKDTFTKNTNS 895

RESULT 3  
 RBP1\_PLAYB  
 ID RBP1\_PLAYB STANDARD; PRT; 2869 AA.  
 AC 000798;

DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.  
 GN RBP1.

OS Plasmodium vivax (strain Belen).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=31273;

RP SEQUENCE FROM N.A.  
 RA MEDLINE=9231538; PubMed=1617731;  
 RX Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;

RT "A reticulocyte-binding protein complex of Plasmodium vivax  
 mezoites";  
 RL Cell 69:1213-1226(1992).

CC - FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO  
 CC HUMAN RETICULOCYTE CELLS.  
 CC - SUBUNIT: HOMODIMER (POTENTIAL).

CC - SUBCELLULAR LOCATION: MEMBRANE-BOUND.  
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

CC EMBL: M88097: AAA29743.1; -  
 CC HSSP: P36956: 1AM9

DR Malaria; Receptor; Signal; Transmembrane.  
 KW SIGNAL 1 17  
 FT CHAIN 18 2869  
 FT DOMAIN 18 2807  
 FT TRANSMEM 2808 2826  
 FT DOMAIN 2827 2869  
 FT SITE 1030 1032  
 FT SITE 2599 2601  
 SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CR364;

Query Match 2.5%; Score 301; DB 1; Length 2869;  
 Best Local Similarity 18.1%; Pred. No. 1e-05;  
 Matches 455; Conservative 343; Mismatches 890; Indels 824; Gaps 108;

OY 12 DEADKHVDFEGQKVDHVEHGRK--NYVELKGSLSLASILGETAFYVKSQ 64  
 DB 675 EEOIKELIDKMAKRYHLYKELLSLKSSSVYFTENNELNTASYNMEEGSAKKEADND 734

QY 65 -----SKYTE-----LIEANSKRNPCCKDG-----KGNVDVRFVSKBOAGYDN 102  
Db 735 INALYNSYREDINALIEBEVEKFEVTENKESTLEMLKDEMEBEKLODAKETAFKALNFVSD 794  
QY 103 K-----KMKCSGMCAPPRRLHLCKNFPNMNSNDSSAKADDLLAEVMAAKYEBESIKT 158  
Db 795 KLTJVTYTMASAVTNAGIEKKEIKELAKOFEN-----HKMKEFSDASTKEALON 845  
QY 159 HYPKDSKYPGSDPFPMCTMLARSPADIGDITRGRLYLGNKKKONGKETEPEKLEOKLK 218  
Db 846 SMOQYNOE-----GDAL-----EKHKONSEKEBEYFKN--- 874  
QY 219 EIEFKIHDLKDEAKRYNGEDENPFYKLRDWMJTANRETVMGAMTCSKELDS----- 273  
Db 875 ---ESVEEDLSRETEBEQYTKHKNNFSRRKGE-----ISAEITMREVIKNIESOL 923  
QY 274 SYFATCNDTQOGPSQTHNKCRCCKDKANAGKPRAGGDVITYPTTYDYVPQYLWFE 333  
Db 924 NY-----GYIEKYFSLIGDNEVSTAKALK-----EKIVSDSLRDKIDQYETEPEKE 970  
QY 334 MAEDFCRKKKKLENLKOCGKDK-----SDEYRYSRNGYDCE-----OTISRK 379  
Db 971 -----KTSAVENYVSTOSLSKAIDSLKRLNGSINCKKYNTDIDLRSKITLREE 1022  
QY 380 GKVRMGKCTDCFPACSGYENMIDNQRKF-DQOKYTKEISDGGGRKRAVGTTKREG 438  
Db 1023 VQKEPKRGDKC-----GEVTTALLKSLRDKGKINELKNDG-----RLNSLDTKKED 1071  
QY 439 YEKSFYER-----LKNDGYVDAFLGLLNNEKACKDITDGKINFEKVSNGGVGGS 493  
Db 1072 LKREYSESKSIHLKSKDQKPODP-----LNRIDEMEDI-----KRDVDELANYQYI--- 1119  
QY 494 GGTSGASGTNDENKGTFRSE---YCOPCPCGVQHK-----GGMOMERTKYVKM 541  
Db 1120 -----SENKVTLEKNNSVTYIE-----AMHSIINTVANGITSNKKEILLKSAVEV 1163  
QY 542 R-----WSKLYKPIKGNKAVLLK-SLKVYKDMILKMKKECLTONSSDGSVG 589  
Db 1164 EDKLNLEVDNEDYKVKVKNPEKQLEAIRGSMKLE--VINKHYSE--MPOLE---STA 1216  
QY 590 SVVTTGASGSGSEK-----KELYDEMCKCYHNEYOKYVNOGEVEED 630  
Db 1217 NTLKSNKKGKENEHDELEANKTKGOMRDYELKRIAELEKTEGYNELKDNENKANKEP 1276  
QY 631 DDELKAGGL---CILPMPKKNKEVSEAKSONNHAD--IOKTFPDFYVVAHMLKDSIH 685  
Db 1277 EPEKNITICHVLERTVEKDKAGKYVEEMNSLKTIKLIQETSDS---ONELVTTSTI- 1331  
QY 686 WTRKRLKSCISDGTMTKCRNGCNKRCDFEKVWVOKETEMKPIKDHFKTOGIEGYEYFT 745  
Db 1332 --THLENAKG-----YEDVIKRNEDSIQLRBKAK----- 1360  
QY 746 TLELILKLOFLKEDTEENTENSL--DAEAEELKHLOKILKLEBNN---LAVYNAGTE 799  
Db 1361 SLETLDEMDKLVQOVNMNLGSAIOGNAGISKLELNEKLVIELLSTNYSILLEYKKNSS 1420  
QY 800 QKTLMCKLNLHENDATCKDCPLPEEDKSRGSRADSPDIFIRPREKEDDENEDDED 859  
Db 1421 ESVAFSQLANOEFTKAE-----GEEKNASARLAEAE---KLKEDIVADLDYSDID 1468  
QY 860 EVRDEETAKETTESATDTTSLDVCPIYGVKVLTKDNESLODACS-----LK 907  
Db 1469 KVKKIEGJIKREILMKESALT-----FWESEKFKOMCOSSIMENAKGKKKIE 1516  
QY 908 YGGNNSRLGMCVTPS-----GEPTTSS-----DKNGALCYPPRRRLRYIKKIY 951  
Db 1517 YLKNGDGCGKANTIDSOMEVEGVNYKAEHAFTVEAQVDDTKAFC-----ESIV 1566  
QY 952 DMATKTES-----POASGSFASSTSGSTTPPOSKKALLAFVESALETFFFL 998  
Db 1567 AYVTKMNLFPNLSMKVEYKCKEKNDEAEKYSAKLKPYDGR---IKRAYSENERKISEL 1623

QY 999 WHRRYKEEKRAVAQ--EGAGHGLPRVEEGSPREYDPEDLKE-GK1PDGFLKQMFYTLGDYR 1055  
Db 1624 KEKAVEKKESSQLNDVSTKSLQIDNCRQOLD--SVLSNIGRVKQNALQ-----YF 1673  
QY 1056 DILFSGSNDTTSVSKDPTSSSNDNLKNVYLLASGSTBEDEBRKMKYKKEI--KNFRKCSTE 1113  
Db 1674 D-----SADKSNKSVLPISLGAELSLDKVAAKESYKRNLETQVNE 1715  
QY 1114 RSAPLVSHPQTMWNNNGKIYHGWVCALTSKDKIAKVEKPKOKEIENPMLDEANKP 1173  
Db 1716 MSRIWVEGSLT-----DIDKKITDIEN-----DLKMK 1744  
QY 1174 KPOYQYTNV-KLDENSGTSPRTTQOASSONT---PTTLHFYKRPYTFKWEWESGF 1229  
Db 1745 K-OYEEGLLOKIKENADKRSNFEVSEINALLDPST-SIFI----- 1785  
QY 1230 CREKKRLQIKVDCKVENGVGKSGDGEACDS---ISTH-----DYSTYPSNCPGCG 1281  
Db 1786 ---KLKLEKXDMTGDLKNYGVKKNELHGEPTKSYNLETHLSNATDYSVT-----F 1833  
QY 1282 KHGSSYRWIERKKIEPHKQSNAYGOOKTDATRNNGNTFDKEFCFCTLETWPDAAKFLERL 1341  
Db 1834 EKAOSLRELAKEEBHLR-----REBEALFLLNDIKYVES-----LKL 1873  
QY 1342 KNGPCKTKKEYGDDIDFEKDKSTFOHTEYCGPCPKFKTCNGCNGVSGINGCDDGKS 1401  
Db 1874 KEMMKVSAEYEG----- 1886  
QY 1402 IDAKEIAKMSSTTDVYKRVSDNDOTNTEFGDDLKDACOHANIFGIRKDVKKCGYGVYD 1461  
Db 1887 -----MKRDHTSVSOLVODMKTYDELKTLNDISCSVLNANNYSIYK----- 1930  
QY 1462 ICEQTNINERTDGEKEYIQRALFKRWENFLEDYKNIKDKISHCIKKGESKJINGCEKN 1521  
Db 1931 VKESKHADYRDANSMTESMTLANY--FLSDAKIS-----SGHEFN 1971  
QY 1522 SKCLEKW-----IEKKIAEMEN--IKKRPENDYENKQDPD-YNVASITL 1561  
Db 1972 AEMKSNFOTDELEFISVYSNSELKKIEBDSNDVIOKREESQALKDARDIYVNIKLK 2031  
QY 1562 EELLPIKIAVNDQNVK-----LCYFENSKGCTLNSNQNKNENDAID 1605  
Db 2032 NEFNEKLEEAKNKEBYSEKVAEALKRLSOYEGIRCHEFNH--RLDNTE-ELEN--- 2084  
QY 1606 CMLKKLGKAKNCGKPSGEGKOSDCKEPPPLPEDEODQPEENTLEPPKFCPTQPREK 1665  
Db 2085 --LKKMTYIYD-----KKSREBSGLQEM---ENEMTYNSI-----TQL 2120  
QY 1666 GGETCGNKEEKKD-EKKEESEPAKESGPAABEPAPTAESSETETNFPPEPGTGAAPP 1724  
Db 2121 EGIYVASAGESKEDIKLEKRSNEMRNIS---EKISTIDSKYIEMN----- 2162  
QY 1725 STRAPPTPDTPPLRPAQDEPDFSTIIQTTIPFGVALALGSIALFLKTKKASVGNLFQ 1784  
Db 2163 -----STIDELYKLGKNCQAHMISLISY-----AMKTSK 2194  
QY 1785 ILQIJKS-----DYDIPLKSSNRYIPIYSDRYKGYIYMEGSDDEDKYAFMSDTT 1836  
Db 2195 LIMINKKEKENCVDYIKDNSSSTDGIVETLKGFYSKULF-----SSASEIYQNMADTY 2249  
QY 1837 DVTSESEYEELD-INDI-----YVPGSPK-----YK 1862  
Db 2250 SVNPAKHEKESLNAIRDIKRELKLFHONSDISIYEGVQNMALAYDKLNEKREMDELR 2309  
QY 1863 TLIIEVLEPSSGNNTA-----SKNTPBDTENDIONGCIIPSSKTTDNEMNOLKKEFI 1914  
Db 2310 NISETKLKQMSHSDVFRPMELIHKGN--ETNNKSLLEKELKLVNDHMHSMEMEMI 2366  
QY 1915 SNMLQONPDVNDYTGNSSTNTNITTSRHNVDNNTNTJMSRDNEE-----NLL 1966  
Db 2367 KNGLKATPEBQON--INNITYIEAEVKTLEBIDIDYGDQNVQIYEBHKKQPSILDRKNAL 2445  
QY 1967 LPSIHGDNLYSGEESYVNVNKNVS-----MNDIPINBDN-----NVYSID 2007

```
Db 2426 MODIE-----IFKKENNYNIMEVNTETIHRVNDYIEKINTKLVQATEYRQLENTKOND 2481
Qy 2008 LI-NDLSGSGKPIDIYDEVLRKE---NELFGTEMTKRTS-----TONVATNTSDPIH 2057
Db 2482 MLONFIFLKVSIIEFYEVNKKKESILNDLYEQERLKLIGEHLDEIKRNVETLSYEID 2541
Qy 2058 NOLELFHMKDHRMCEMKKEDILN-----KIKEENKENINNSGTYSDNKPSSH 2112
Db 2542 OKMEMSK-----MLER---KSKMNTSYLELERANE--INNDAKQIKDDDT--- 2586
Qy 2113 HVLNVDVSIQIDMDNPKTKNETTNDKSTMDTIILDELEKYNDPYD-----FYED 2167
Db 2587 -ILNSVLEAI-----OKRSDMAIFSQMADRNPNETKYSAEKYMMA 2628
Qy 2168 DIIVHDVDEKSSMDIYVDHNNVTSSNMNDVPTKMHIEKNINVKKEIFE 2219
Db 2629 NEIIRQLVEKLRIGQLVQDSISLSE-----MNSKKSALIEKE 2666

RESULT 4
Ym67_YEAST STANDARD: PR1: 1658 AA.
AC Q03661; Q04988;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE HYPOHECTICAL 187.1 KDA PROTEIN IN GDA1-ERG8 INTERGENIC REGION.
GN YMR219W OR YMR261.13 OR YMR959.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-711 FROM N.A.
RC STRAIN-S288C / AB972;
RA Deidman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
RL Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 608-1648 FROM N.A.
RC STRAIN-S288C / AB972;
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
CC EMBL: 249809; CAA89934.1;
DR EMBL: 249939; CAA90190.1;
DR SGP: S0004832; YMR219W.
KM Hypothetical protein.
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match 2.3%; Score 277.5; DB 1; Length 1658;
Best Local Similarity 18.0%; Pred. No. 5,6e-05;
Matches 344; Conservative 290; Mismatches 670; Indels 605; Gaps 89;

Qy 601 SEKKELYDEMKCYKHNEVQKVVQGEV-----EDDELKGAAGLCILPNPK----- 647
Db 2 SKKKEFFP-----RANKLKLTPRRKLKLTSLDLADESKKMDQGYSRVINDKRYRAK 57
Qy 648 -----KKEVSEAKSONNHADIQKTFHDFEYVVAHMLKDS---IHWFTKRLKSCISDG 698
Db 58 PTOHSTLHESISSRRSSHIT---NKSLSH-----DSARALSW---VDSLINGR 99
Qy 699 KIMKRCNGCNKCKDCFEKVVVKQKTEPMK-----PINDH----- 731
```

```
Db 100 KSI--LITLEKEDALFERKLEERQRQFQJLHDSLMNKYTNQNSHQRLLI LRKSQYGTDS 157
Qy 732 FKTOEGIP-EGYFFTTLELILKLOFLKEDTEENTENSLOAEAEELKH QKILK---LEN 787
Db 158 FQNDIEPLDSFISPL-----PDAEDESSNIDSDKDEDLEGR QSLIKDFDEN 207
Qy 788 E-----NMLAVNAG 797
Db 208 DEYELESEERKNSDGOSSPIMILSDEYAEAGLODVSNDEYAEBOY ERKNIGOEQAN 267
Qy 798 TEOKTLMKDLNHELNDARKKDCPL-----PEEDKSRG-----ISAD-----PS 837
Db 268 VENAFOISSDSSSEGOINSSEGVEMELIEDDIDYSDAEKESOGABTEI:VUPFSKYMQR 327
Qy 838 PD-IFPRPEEKED-----NED-----DOEDEVDNDEIFAKETEGSAT 877
Db 328 TUNTKIPVIEKYESDEHKVHQRYSEGDARDFGSVNISYDDESE---DEISQAEYSANAE 384
Qy 878 D-----TTISLDVCPYVKGKYLTKDNESLDQACSLKYGNN-----SRGMICVTPSGEPTTS 929
Db 385 NYVHNHEHLDKRELIEDIESDSSES-OSAQSEDEGSEDFEYKMKNEI:STSEETNTSE 443
Qy 930 SPKNGAICVPRRRRLYIKIYDMAATKTESPOASGEASSTS-----GS:TPPDSKPAALK 985
Db 444 SRDQGF-----AKDAYTKNKVEQOENDEPEKEDIIRSLDKNFRGININKSEYSENYLE 497
Qy 986 AIVESAIEFTFLMHRKEEKAQAEOGAGHGLPRVEEGSPEDPED----- 1032
Db 498 NETDPAIVE-----RENGINVEGYDVYKGSVESLDLHESPDNL:DLAARAMLQRO 548
Qy 1033 -----KKEEKIPDGF:ROMFYTLGDRDLFGSSN-----D:TVSKDTPSSS 1076
Db 549 OSRNSNCPQKEOVSESYLGHSS-----NGSNLSGRSLDESEI:QILPKDFTGEN 596
Qy 1077 NDNLEKIVLLASGSTFOEEREKNKKYKEIKNFRKCSYTERGAPMLVSHQ:VWENNGKIYH 1136
Db 597 NNNLKTDRDLDSSVLEIEVEKYSKKLGSTKEKLYPLSTDTTINNSS:GANEISTY 653
Qy 1137 GMVCAITSKDIKAGVEKRP-----QKLENENLMD-----EANKI:PKRPQOYQTVN 1183
Db 654 -----SLDDADALISENTDVP:PLMEIKTPKYEYVVISBSYSGSYEDNT:AMPPOVEYTSR 709
Qy 1184 KIDENSGTSPRTTQOASDNTPTTILTHFVKRPTPRMVEEM-----GESEF 1229
Db 710 FM--NDPFSNLDDYEKKRDLKSTLALA--PAFTKDAEVEAGVT:ISCLTSTSGHYN 765
Qy 1230 CREKKRRLKQIK-VQCKVNGDVGRCSGDGEACDSISTHDYSGVPSFM:PGGKHCSYR 1288
Db 766 IHTTSKETKQVSDLDDESTEN-----VFENNENTDEKKNOSKN:PGV---ANSTD 812
Qy 1289 KWIERKIEFHROSNAVYGOOKTDATRNNGNTPEKCEKTLTETWPAK:TLERLKN----- 1343
Db 813 KSTEDNTDEKYSALNY-----TNVTGDSSECDIETIASVNEERL:CEKDMNEAEMS 865
Qy 1344 --GPC-KTKKEYGGDDIDEKXS-KTFQ---HTEYCGCPKFKTKQ:NGNCVGSJLKN 1395
Db 866 SGPDECYKQNDPSKTIQISFTSDSPDNFQESNDNTERSS--TRYKVR-----NSD 912
Qy 1396 CGDGSIDAKELAKMSSSTDVVMYRSNDT--NTEFGDDL-----KDA:QHANIIFGIRK 1449
Db 913 LEBDESL-KKELTK-----AEYVADKLDEESEDSTEDYADAPGND:SENENIYAKGIRK 966
Qy 1450 DVMKCGYCVGVICQETINERTDGEKEYIQIRALFRWVENFLEDYNK INDRISHCIKRG 1509
Db 967 DE-----LGIVEPEN-----EKVKNK:VHEE--ETLFPEA 991
Qy 1510 ESKSCINGCEKSKKLEKIEKIAEMENIKRFPNDQYENKQPOPYNVN:SLLELIPKIA 1569
Db 992 NVSSSVN-----YONKDMHTDVINOEAQANAEAGERKYYIQ:TWDEE---AHIS 1037
Qy 1570 VYNDQDNVILKLVFENSKCCTIISNTQNNKENDALIDCM:KLIGVAKN:PGRPSGKQSD 1629
```

D	b	1038	IIERID-----ENAIg-----NNNE-----IPERSVEKTHNEVLE	1065
Oy	1630	CKEPPLPEDEEDONPENLTPPEPCFPPTOPPEEKGETCGNKEEKKDEESEEPak	1689	
D	b	1070	RR-----ATTENTKALENNTN-MHDVSQACSDSDRDODSTAEKVE---	1111
Oy	1690	EESGPAEEPAFTASETETNPPEPGCPAPAPSTPAAPPDPPRPPADAPDPST	1749	
D	b	1112	---GSAHNLDIVSSSEIE-----SVEPKPSDR--SN	1141
Oy	1750	ILQTTFPGVALAG-----SIAFLFK---KTASVGNLFQIQLQIPKSDIFTLKS	1801	
D	b	1142	IESSPINAVGVAVGVAKYVDVAESPVKIKDVWDSSDON-----VDIGVINQIIFKS	1195
Oy	1802	NRYTPVSDRYKGKTYTYMEGDSDEDKYAFAWSDTDTVTYSSESLEYELINDIYVGSPKY	1861	
D	b	1196	NSTDASYN---MKSVSKERDSEDDEAVILGGVT---AEA#-----ND-----	1232
Oy	1862	KTLEVVLPSGNNTTASGKNTSDTRNDIQNG-IPSSKITDNEMNLQKKEFISNMQN	1920	
D	b	1233	-----NGNSRYI-NIDPTNGAIEEDESEVFROOVKDKENLHKSEPLEVGLQS	1280
Oy	1921	QPNVPDYTSGNSTNTITTSRHVNDDNTMTSRMMEENLLPIHNGNLYS---	1977	
D	b	1281	EQHEKKDHSHNEEFOTIYGDTITSANISHN-----APDDIKOOLLKLMSLENYSORL	1333
Oy	1978	-----GEETYNVAMVMAMDIPINRDNVYSGIDLINDLSGGRPIDYDEV--L	2026	
D	b	1336	IEDSRKRGKNGESEDVN-TSEREDLFFEPS-----VNEKYAYGIEDPFSELDISI	1385
Oy	2027	KRKENE--LEGTEMTKRTSQONAKTTNSPIHNOLEFHKLMDHRDMCEWKAKEDIL	2084	
D	b	1386	QHAPHEDDLDSNQERS-----IEELNSEP--EEALEYLEIEGTETAASSKMND-	1436
Oy	2085	NKLEENN-----KENINNCKTYNSDAKPSHHNVLMTDVSLQIDMPNPKTK	2131	
D	b	1437	--EORGINISTDLPSPBPDKEEVTDSPYSNSEITTAKSAPSPSEYELFTSTP---	1491
Oy	2132	NETNMDTN-----ODKSTMDTIIDLEKRYNDPYDFYEDDIHYHDVYE-KSS	2180	
D	b	1492	NEVP-MEINDEIPATLTLEKHDKTNVSLVDREH-----LASHVDVNEPHDN	1538
Oy	2181	MDDIYVHNNTVTSNN-MDVTKKMH-I-EMLVNNKKELFEETPISDIYN	2227	
D	b	1539	SINKVNEGEPERHOAVDIPVKYKEVEQEDEMPSKVLEEOKPSMELIN	1587
<hr/>				
R	sULT_5	MSPL_PLAFK	STANDARD;	PRT; 1630 AA.
A	c	PO4932;		
R	x	13-AUG-1987 (Rel. 05, Created)		
R	x	01-FEB-1996 (Rel. 33, Last sequence update)		
R	a	MacKay M., Goman N., Bone N., Hyde J.E., Scalfe J., Certa U.,		
R	t	Stuenkelberg H., Bujard H.;		
R	t	"Polymorphism of the precursor for the major surface antigens of		
R	t	Plasmodium falciparum merozoites: studies at the genetic level.";		
R	t	EMBO J. 4:3823-3829(1985).		
R	n	[2]		
R	p	REVISIONS, SEQUENCE FROM N.A.		
R	a	Pan W., Tolle R., Bujard H.;		
R	L	Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.		
R	c	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR		

Query Match	Best local similarity	2.2%: Score 269; DB 1; Length 1630;
Matches 264; Conservative 190; Mismatches 445; Indels 552; Gaps 60;		
Qy 978 DSKKALNAF--VESAIETFFLMHRYK-----EE--KKAVAGQCA-----	1014	
Db 23 ESYQELVKKLELDAVAVLTGYSLFHKEMILNEEITTTGASAGSGTSGTSGSPGS	82	
Qy 1015 -----GHGLPRVEGSEYDPED-----KLEKGIIP-----	1040	
Db 83 GTPSSRSNTLPRSNITSSGASPPAASDASDVADLKHVRVNYLTITIKELYPOLFDT	142	
Qy 1041 -----DGF--LRQFYVLTGIDYRDLITSGSDNTS-----VGKDPIS	1074	
Db 143 NHMLTLCDNIHGFKLIDYEEINELLYKNTFFDLRAKLMDVANCYCOIFPMLKIRA	202	
Qy 1075 SSNDNLKNIIVLLASGSTQERE--KMKY-----KEIKNFKCSTERSAPNLVSHPOTW	1126	
Db 203 NELDVILKLVFGYRKPLDIKNVGMEDYIKKNKKTITENINEL--IESKKTIDKKNKAT	261	
Qy 1127 WENNGKITYT-----HGAVCALTSKDKTAKGYEKRPKQIENENIMLDEAN	1170	
Db 262 KEERKKKLYQAOYDLSIYNKOLEEAAHNLISYLEKRIIDLTK-----KNENIKELLDKIN	314	
Qy 1171 KKRPQRYQYTWKLDKENSSTPRITQYQASSDNPPTTLTHFVKRRPYFRFMEEGESPC	1230	
Db 315 ETKNPP-----PANSGNTPNLLDKNK-----I	338	
Qy 1231 RERKRLKQI-----KYDC-----KVENGDYGRCSGDSEACDSISTHDYS	1270	
Db 339 EHEKEIKELIAATIKFNIDSLFTDLELEYLREKKNKIDI---SAKYETKSTEPNEYP	395	
Qy 1271 TVPSFNCPCGCHGSSYRKWIETKKIEPHKOSNAGVQOKTDTATRNNGTPEDECKFTLET	1330	

```

Db 396 NGVTV-----PLSYNDINNALNELNSFGDILNPDYTKEPSKNIT 436
OY 1331 WPDAAKFLERLKNGPCKTNKEYGGDIDIEKSKTFPOTTEYCGPCPKRTNCGNCGVS 1390
Db 437 DNERKFFINEIE-KIKIEKK-----KLESDEKSYE-----466
OY 1391 GLNGNCGDGRSID--AKEIAKMRSSSTJDDVMVRSDN-DTNFEGDGLK-----DACOHA 1441
Db 467 -----DMSKSLNDITTEKEYEKLINETID--SKFNNDILJTFNEKMMGKRYSKVEKLTJH 518
OY 1442 NIFKGIKRDVMKCGYVCGVDICEOTNINERTDGEYIOIRALFKRWENFLEDYKNKIND 1501
Db 519 NTFASYEN-----SKHNLEKLTALKALYMEDYSLRNIVVEKELKYNNLSJK 564
OY 1502 ISHCIRKGBGSKCINGCEKSNKCLKWLIEKTAEMENIKRRNDOTENKDDPDYVNSKL 1561
Db 565 I-----EMEIEETLVENIKKDE--BOLFEKKITKDN-----KDEKILEVSDIV 606
OY 1562 EELIPKIAVNDQDNVIKLCYFENSKGCTLIJNTQNNKENDALDMLKKIKYKAKNCPCK 1621
Db 607 KYOVOKVILMKIDE-----LKKYQILILNVEIK 635
OY 1622 -----PSEKOSDCKEP--PLPDEEDQNPENTLEPPKCPPTTQPEEKGETCGNK 1673
Db 636 HNHHVPSNYKQENKQEPYLIYLVKKEID-----KLKVFMPKV-----ESLINE 678
OY 1674 EKK--KDEKKESEEPKAE-----ESGPAEERAPYAESEETENFPEPGET 1718
Db 679 EKKNKITBEOQSNSEPSREGELTGOATKPGQAGSALLEGSDVOAOEOKAOOPVPPVP 738
OY 1719 GRAPPEPPAPPTPTPLRQADPEPDSITLQTTIPGVALAGSIAFLKKTXTAS 1778
Db 739 VPEAKAQPPTPPAP-----VNKTE--758
OY 1779 VGNLPQIOLIPKSDYDIPLKSSNRYIPVSDRYKGYIYMEGSDDEKXAFMSDITDV 1838
Db 759 -----NVSKLDY-----LEKLYEFLNYSYICHKYLIVSHSPTMANK--ILMOYKLT 801
OY 1839 TSSESEFEELINDIYVPGSPKYLIEVLEPPSGNNNTTASGNTPSDTRANDIONDGP- 1897
Db 802 KEESKLSKSCD-----PLDLLEFNIQNN-IPV 826
OY 1898 ---SKKITDNEMNOL-----KKEFTSNMLQNPNDVNDYTSNGSNSTNNTTSHAND 1949
Db 827 MYSMFDLSLNNLSQLPMEIYERKEMVCMYKLDNDIKILLEAKKVSYSVTLSSSSMQ 886
OY 1950 NNTNTTMSRDNMENMLLPISHDGNLYSGEYSYNNVNVNSMNDIPINRDNVYSGIDL 2009
Db 887 PLSLTRPODKPEVSAN-----DQTSHTNINLSLK-----LF 917
OY 2010 NDLISGKRPIDYDEVLRKKNELFGTENTKRTST-----ONVAKTNSDPIHOLELF 2063
Db 918 ENILSLIGKKNKYOEILIGKSSSENFYEKILKDSDFYNESEFNFYKSKADD--968
OY 2064 HMLDRHRDMCKMKNKEDILNKKEE-----NNK-----ENINNSGKT--2102
Db 969 ---INSLNDESKRKKLEBDI-NKLKTYLOLSFDLYNKYKLERLFDKKRYGKKYKMOJK 1024
OY 2103 -----YNSDNKPSHNHL-NTDVSIOIDMDNFKTNEITNMOTNODKSTMDTI 2149
Db 1025 KTLTLKEQLESKLNGLNPN--KHVLONFSVFF-----NKKKEAEI-----ETENTLE--1070
OY 2150 LDDLEKYNDPYIYFEDDIYHDVD---VERKSMDDIYVDHNNVTSNMADVPYTKMH--I 2204
Db 1071 -----NKKILKHKYGLVYKYNNGESSPLKTLSESIQTEDNVASLENFKYLSKLEGL 1123
OY 2205 EAMNIVNKKKEI 2215
Db 1124 KDNLMLEKKKL 1134

```

```

ID MSPL_PLAFW STANDARD: PRT: 1639 AA.
AC P04933:
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMSA) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Wellcome).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5848;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=66014355; PubMed=2995820;
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., R-veros-Moreno V.,
RA Nicholas S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA Freeman R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum merozoites."
RL Nature 317:270-273(1985).
RN [2]
RP REVISIONS.
RA Holder A.A.;
RL Submitted (MAR-1991) to the EMBL/Genbank/DBJ databases.
CC -! SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -! P1M: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL, X02919; CAA26676.1; -.
DR PIR, A24594; A24594.
DR InterPro: IPR000561; -.
DR Pfam: PF00006; EGF 1.
DR Malaria: Merozoite; Polypeptide; Repeat; signal; Glycoprotein;
KW Transmembrane; GPI anchor.
FT SIGNAL 1..19
FT CHAIN 20..1639
FT CARBOHYD 116..116 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 268..268 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 764..764 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 768..768 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 783..783 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 844..844 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 920..920 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 964..964 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 1058..1058 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 1165..1165 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 1174..1174 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 1445..1445 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 1526..1526 N-LINKED (GLCNAc...) (POTENTIAL).
SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C876E CR-c54;

```

Query Match 2.2%; Score 265.5; DB 1; Length 1639;  
 Best Local Similarity 18.1%; Pred. No. 0.00019;  
 Matches 260; Conservative 189; Mismatches 457; Indels 533; Gaps 58;

```

OY 954 ARTTESPQASGEASSTSGTTPPDSEKALLKAFVSAIEFFFLMHRKKEKKAVADQG 1013
Db 61 AVTTSTPGSKGSVAGSGSVASGVSASG-----SVASGG 104
OY 1014 AGHGLPRVEEGSPEDYDEPK-----LKEGKIP-----1040
Db 105 SGNS-RRTPNPSNDSDAKSTADLKHRVRYNLITIKELKYQFLDINMHLTLDNINH 163

```





```
FT CAROHND 239 239 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CAROHND 470 470 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CAROHND 536 536 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CAROHND 607 607 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CAROHND 802 802 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CAROHND 899 899 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CAROHND 919 919 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CAROHND 965 965 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CAROHND 991 991 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CAROHND 1089 1089 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CAROHND 1196 1196 N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT CAROHND 1588 1588 N-LINKED (GLCNAC . . .) (POTENTIAL) .
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF6EA98 CRC64;

Query Match 2.2%; Score 265; DB 1; Length 1701;
Best Local Similarity 19.1%; Pred. No. 0.0002;
Matches 358; Conservative 241; Mismatches 659; Indels 620; Gaps 90;

QY 565 KDMMLKKMKKEFCCLNONGSGSVYTTG-----ASGNSKKKLYDEMKCYKHN 616
DB 48 KKKMVLNBESTGTAVT-ISTPGSGSVTSGSVASVYASGSG----- 91
QY 617 EYQKYNVOGEVEEDDELKAGGLCILPMPKKNKEVSEAKSONNHADIQTFHDEFFYY-- 674
DB 92 -----GSV-----ASGSGNSRRTNPSSSDSNKT--YADLKHRYQNTLFTIK 134
QY 675 -----WVAHMLKDSIHMTKRLKSCIS-----DGTKMCRNGC-N 708
DB 135 ELKYELEFDLTJHML--TLKSNVDFKYLIDGYEELINELLYKLNFYDDLRAKLADACAN 192
QY 709 KKCDC-FEKWVQKETE-----WKPIKHFKTOBESIPRGYFTTLELILKLOFLKE 758
DB 193 SYCQIPNLIKIANELEDYKKTIVFGYRKPL-DNIKDNVGMEDYIKKNTTIANINELLE 251
QY 759 DTEENTENSIDAEEABELKHLQIKLENNENLVAVNAGTEQKTLMDKLINHELDATKC 818
DB 252 GSKKTLIDQKNKNDNEGGKKL-----YQAOYMLFIYNKQLOBAHNLISLEKRIDILKKN 306
QY 819 KCCP--LPEEDSKRGSRADPS---PDIFIPRPEEKEDDENEDDEVEDDETAKEETT 872
DB 307 EIKKILLEDIDIKITDAENPTTGSKPN--PLPENK-----KKEVEGHEEKKE-- 352
QY 873 EESANDTTSILVCPILYGVVLTKDNESL-QDACSLKY--GNNSRLLGMCVPPSGEPTS 929
DB 353 -----IATYIKFNIDSLFTDPLELEYLRKNNKVD--VTPKSDDPK 393
QY 930 SDK-----NGAICVPRRRRLYIKIYDMATKTESPOAGSEASTSGSTTPPDSEKA 982
DB 394 SVOIKRVPFNPNGIYVLP-----LTDIHNSLAADNDKNS-YGDLANPDPTKEK 439
QY 983 LKAIVESALETFFL-----WHRYKKEKKAVAOAGAGHGLPRVEEGSPEDY 1029
DB 440 INEKIITDOKERKIFINIKQIDLEEKNIHNTKEONKLLD-----YEKSKDYE 491
QY 1030 P-EDLKAKEKIDGFLRQMFYTLGDIYRDLFLFSGSNDTYSK-----DTPESSNDNLKNT- 1083
DB 492 ELLEKFEYEMKFNNDK-----DVVDKIFS-ARTYVVEORVYNNKSSNSNSYVNO 543
QY 1084 -----VLLASGTEOREKMKNYKEIKNFKKCTERSAPRLVHPOTMMENNG 1132
DB 544 KKKALSTYEDSLRKGISEKD---FNHYTYLKTGLLEADIKKLTETIKSENKILEKNK 600
QY 1133 YIMHGWCALTSKDKIAKVER-----KPKOIE-----NPNLMDPEANK 1172
DB 601 GLTHSANASLESDYIKLOVQVLLIKLIEDLRKTELFLKKNQKDSIHVPNIYKQNK- 659
QY 1173 KRPQOYQYINV-----KLDE----- 1187
DB 660 ---PEPYLIIVLAKKEVDLKEFIIPVKMDLKKQOAVLSSITQPLVAASETTEDGSHSTHT 716
QY 1188 -NSGTSPTTQOASDWTPTTLTFHFVVRPTYFVFMFEEMGSGFCERKKRLKQIKVDCK 1245
DB 1188 ---NSGTSPTTQOASDWTPTTLTFHFVVRPTYFVFMFEEMGSGFCERKKRLKQIKVDCK 1245
```

```
DB 717 LSGSGTEVEETEVEETEVTGHTTYYTLTP-----KEHSAPKEVKN-- 759
QY 1246 VENGVDGRSGGCEA-----CDSISTHDYSTVPSNCGCGKHCISYKNIERKRI 1296
DB 760 VENSIEHKSNDNSQALTKTVYLLKLEDFLTKSyi-----CHKYIIVSNSSMDOKLL 810
QY 1297 E-----FHKQSN-----AYGOQKTDATRNNGHTPFKE--FCK 1326
DB 811 EYVNLTPREKERLKSQDPLDLFNQNNIPAMYSLYDSNNNDLQHLFPLYKEMHYIYH 870
QY 1327 TLETMPDAKFLERLK--NGPCKTKEXGGDIDFEKOSKTFQHTGYCYPCPKFTNCON 1384
DB 871 KLKEENHIKKLEEQKQITGSTSSP-GNTTVN-----TQSAATH-----SNSNG 915
QY 1385 GNCVSYGLNGCDDGKSIDAKELAKRSSTTVYVARYSDNDNTFEGDYLKACQAHANF 1444
DB 916 QOSNMS-----STNTONGVAVSSGPAVVEESHJPLTVLSISNDL 954
QY 1445 KGLRKDVWKGYVCGVDICEQININERTDGKEYIOLRAL-FKRWENF-JEDYKN-INDKI 1502
DB 955 KG-----IYSLNLGKTKVNPPLTISTTEMEKFEYENI:KANDTYFNDDI 999
QY 1503 SHCIRKGGESKCIING--CEKNS-----KCEKWIETKKIAEMENIKKR:NDQYENK--D 1551
DB 1000 KQFVK--SNSKYITGTETQKNALNDEIKKLKDPLQLSFDLYNKYKLLDLFLNKKKELG 1057
QY 1552 QPDYVAKSLT---EBELPKIAVNDQDNVYK--LCYFENSGCTGLISNTJNKNENDAIQCM 1607
DB 1058 QDKMOIKKLTLLKEOLESKLNSLNPNHVNQNFSEFNNKKEAEIAETENTLENTKI--L 1115
QY 1608 LKKIGYKAKNCPKPSGEKQOSCKEPPPLPDEEDQNPENTLEPPKFCPTQPREKGG 1667
DB 1116 LKHGYGLVYVYGESS-----PLKTLSEVSIOTEDNYANLEKF-----FALSIDIG 1161
QY 1668 ETCGNKE-EKK-----DEKKESEEPAAKEESGPAEPAFPAESEETEPTNP 1713
DB 1162 KLNNDLHLGKKKLLFSSGLHLHLELKEKVIKKNKNTGNSPSENNKKI NEALKSYENFP 1221
QY 1714 EPPGCGPAAPSTPAFPPTD-TPPPLRPOADEPFDSTILQTTIIPRGV:LAUSIAFLFK 1772
DB 1222 E-----AKVTIVVPPPOGDPVTPSP:SVVSGSGSTKEETOIP-----TSSSL 1264
QY 1773 KKTASVGNLFQILOPKSDYDIPPLKSSNRYIPIVSDRYKKTYYIYEGDS-DEDKY-- 1829
DB 1265 -----LTELQOVQLONTDEEDSL-----VLP:FGESBENDEYID 1301
QY 1830 -----AFMSDTJD--VTSESEYE-----ELDINDIYVPG 1857
DB 1302 QVVTGEAISVTMDNLSGFENEYDIYLLKPLAGVYRSLLKQIEKNTI:ENLMLNDILNSR 1361
QY 1858 SPKYKTLIEV-----VLEPSGNNTTASGKNT-----PSDTRNDIQ- 1892
DB 1362 LKKRYKFLDVLSDLMQFKHISNEYIIEDSFKLINSBQKNKTLKSY:YIKESVEDNDIKF 1421
QY 1893 -NDGIP-SSKIIDNENQOK--KEFISNMLQONPNDVPNDYTSNGSNFTNTITTSRHNV 1948
DB 1422 ADEGSIYEVKYLAKYKDDLESIRKVIKEKEKFPSSP--TPPSAPKTD----- 1469
QY 1949 DNNVTYTMRSRDMNEENLPLPSIHGCLNYSGEESYVNVNVMNDPIIN-----BDNNV- 2002
DB 1470 -----DQKESKEFLPPL-----TWIETLVN-NLYVKNIDOLYLAKKAKINDCMBE 1512
QY 2003 ---YSGIDLINDSLSGKPIDIYD-----EVLKRENEFLFSTENTKRTS-----T 2044
DB 1513 KDEAHVKITKLSDLKAIDKIDLFKNTDNFEAIKLLIN-----DDPKKMLGKLSTGLV 1567
QY 2045 QNVAKTTNSDPIHNOLELFHKKLDNRHRCCEKMKKND-----IINKIKE 2089
DB 1568 QNFPNTIISKLEGG--FQDMLNISQHCQVKKQCPENGSGCFRHLDERECKCLIN-YKQ 1623
QY 2090 ENNKENINSGTKYNSDN 2107
DB 1624 EGDK-CVENPNPTCNENN 1640
```

```
RESULT 8
MSPL_PLAFR STANDARD: PRT: 1726 AA.
AC P50495;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (GPI95).
GN MSP-1.
OS Plasmodium falciparum (Isolate Palo Alto / Uganda).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57270;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89005525; PubMed=3049134;
RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
RA Siddiqui W.A.;
RT "Plasmodium falciparum: gene structure and hydropathy profile of the
RT major merozoite surface antigen (gpi95) of the Uganda-Palo Alto
RT isolate."
RL Exp. Parasitol. 67:1-11(1988).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M37213; AAA29611.1;
CC InterPro: IPR000561;
CC Pfam: PF00008; EGF: 1.
CC Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
CC Transmembrane; GPI-anchor.
CC SIGNAL 1 19
CC CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
CC CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 1726 AA; 196174 MW; 5859CEFA2F9A026 CRC64;
Query Match 2.28; Score 264.5; DB 1; Length 1726;
Best Local Similarity 19.08; Pred. No. 0.00022;
Matches 370; Conservative 243; Mismatches 704; Indels 633; Gaps 87;
```

```
QY 561 LKVVKDMILKNNKRCCLPTONSSDGSVGVYTTGASG--GNSEKKELYDEMKCYKHNEV 618
DB -----
QY 98 -----ASGTSGTSASGTSIGTSGTSPSS-----RSNTL 129
DB -----
QY 619 QKAVNGEVEEDDELKAGGLCILPNPKKKEVSEAKSONNHADIOKTFHDEFFYY---- 674
DB -----
QY 130 PRSNTS-----SGA-----SPPADASSDAKS---YADLKHRYVANYLFTIKEL 169
DB -----
QY 675 -----WVAHMLK--DSIH-----WTRKRLKSGISGKT 700
DB -----
QY 170 KYPELFDLTNMLTLCODNTHGFKYLDIGYEINELLYKLFEPDLAKLNDVAND--- 226
DB -----
QY 701 MCRNCGNKKKCCDFEKWKOKET--EMKPIKHFKTOEGIPGEYFTTLELILKLOFLKE 758
DB -----
QY 227 -YCIQIFNLKIRAMELDVLLKLVGYRKL-DNLKDNVGMKEDYIKNNKTTIANINELLE 284
DB -----
QY 759 DTEENTENSIDAEAEELKHLQKTLKENENNLAVVNGTEQKTLMDKLNLHEINDATKC 818
DB -----
QY 285 GSKRTIDQNNKADNDEGKKL-----YQAOYDLISYKKOLEAHNLIISYLEKRIDTLK- 338
DB -----
QY 819 KDCPLPEEDKRGSRADSPDIFIPREKEDENE-DDDEDEVRDDETAKEETTESAT 877
DB -----
QY 339 -----NENIKELLDKINETKNPPRANGTPTNTLDDNKKKIEHEDEKLE----- 383
DB -----
QY 878 DTTSLDVCPIYGVKVLTKDNESL-ODACSLKY--GGNNSRLGMRVYPSGEPTTSSDK-- 932
DB -----
QY 384 -----IATIKFNISDFTDPLELEYLREKNNKYD--VYPKSDPTKSYQIP 429
DB -----
QY 933 -----NGAICVPPRRRLYIKYIDMATKTESPOASSSEASSTGTPPDSCALLKAF 987
DB -----
QY 430 KVPYPNGIVYLP-----LTDIHNSLAANDKKS-VGDLMPNDPKKIKIEKI 475
DB -----
QY 988 VESAIEIEFFLMHRYK-----EKKKA-VAOEGAGHLPRVERGSEPYDP-EDKLEKGRIPD 1041
DB -----
QY 476 ITDKKERKIPFNINIKQIDLEPKKINHTKQNKLLDEYKSKDYELLEKEFEMKFN 535
DB -----
QY 1442 GFLRQMEYTLGDYDILFSGSNDTYSK---DTPSSNDNLKNT-----VL 1085
DB -----
QY 536 NFDK-----DVVDKIFS-ARYTYNVEKQRYNNKFPSSNSYVYVQKALKSLYLDYS 587
DB -----
QY 1086 LASGSTDQERKMKYKIEKFRKCSSTERSAPNLVSHPTWENNGYIWHGMYCALTSK 1145
DB -----
QY 588 LRKGISERD--FNHYTTLKGLADIKLLEIKSEENKILEKFNGLTHSANASLEVY 644
DB -----
QY 1146 DKIKGVK-----KPKIE-----NENLDEANKPKRPOVOYTAV-- 1183
DB -----
QY 645 DIVLQVQKVLILKIEDLKIEFLKNAQLKDSIHVPIYKPNK---PEPYLLVLK 700
DB -----
QY 1184 ---KLDE-----NSGTSPTTQTOASSDNTPTTLTHFYKRPYFRWFE 1223
DB -----
QY 701 KEVDKLKEFIEKVDMLKKEQAVLSITQPLVAASETTEDGHSHTH-----LS 749
DB -----
QY 1224 EWGSEFCREKRL-----KQIVDCKVENGVDGRSGDEA----- 1260
DB -----
QY 750 QSGTEVTEETEETEVGHTTITLPRKEVY---VENSIBKNSDSQALTQTVYL 806
DB -----
QY 1261 ---CDSISHTDSTVPSNPGCGKHCSSYRKIRKIE----- 1297
DB -----
QY 807 KKLDEFLTKSYI-----CHKYIIVSNSSMOKLLEYVNLTPREENELKSCDPLDL 857
DB -----
QY 1298 FHKQSN-----AYGOQTDATRNNGTTFDKE--FCKTLETPDAKFLERLK--NGPC 1346
DB -----
QY 858 FNIGNNIPAMYSLVDSMNNDLQHLFFELYQKEMLYLHLKLEENHILKLEEQOQITGTS 917
DB -----
QY 1347 KTNKEYGGDIDFEKDKSTQHTQTEYCGPCRFKTCNCGNGCVSGLNGCDGKSIDAKE 1406
DB -----
QY 918 STSSP-GNTTVN-----TAQSATH-----SNSONQSNAS----- 946
DB -----
QY 1407 IAKRRSFTDVVVRVNSDNTFTFGDDLKDACQANJFKGIRKQVWKCQYVCGVDICEQT 1466
DB -----
QY 947 -----STNTONGAVAVSSGPAVERSHDPLVYLSISNLDK-----LYSL 986
DB -----
QY 1467 NINERTDGKEYIQTIRAL-FRWVENPLEDYNK-INDKISHCIRKKGESKING---CEKN 1521
```

```

Db 987 NLGNKTVNPPLTISTTEMEKEFYENILKNNDYFNDLIKQFVK--SNSKVITGLNETQKN 1044
OY 1522 S-----KLEKRIEKKIAEMENIKRRFNDQYENK--DQPDYVKSIL---BELIPKIAV 1570
Db 1045 ALNDEIKKIKLKDYLQSLFDLYNNYKRLKLDRLFNKKKELGODKKQIKRLTLKQLSEKLS 1104
OY 1571 VINDOONVIR-LGVFENSGCCLISNTONNKNENDAIIDCMKKIGVAKAKNOGPGPSEKOSD 1629
Db 1105 LNNPNVNLONSVPFNKKKEAEIAGTENTLENTKI--LTKHYKGLVKYYNGESS----- 1156
OY 1630 CKPEPLPDEEDONPEENTLEPPKFCPTTQPPKGGETCCNKE-----EKKDEK 1680
Db 1157 ---PLKLTSEVSIOTEDNVANLEKFRVLISKIDKLNNDHLKKKLSFLSSGLHQLITEL 1213
OY 1681 KESEEPKAESEGPAAEAPPAESETETNPEPPGTCGPAAPPTPADPTPD-TTPPLK 1739
Db 1214 KEVINKNYTGNSPSENNKKVNEALKSYENFLPE-----AKVTYVTPPQPDVTPSPLS 1267
OY 1740 PQADPPDSTLIQTTIPPGVALAGSIAFLFKKTKKASVGNLPOILOIPKSDYDIPTLK 1799
Db 1268 VAVSGSSSGSTKEETOIP-----TSGSL-----LTLELOOVVOLQNTDEEDDSL- 1309
OY 1800 SSNRYIPIYVSDRYKGGTYIYMEDS--DEDKY-----AFMSDTTD--VTSESEYE-- 1846
Db 1310 -----VVLPIGSESDNDEYLDQYVYTGALISVTMDNLISGFENEYDVIY 1353
OY 1847 -----ELDINDIYVSPKRYKTLIEV-----V 1868
Db 1354 LKPLAGVYRSKTKQIEKNIFTEFNLMNLNLSRLKKRKYFLDVLSDLMQFKHISNEYI 1413
OY 1869 LEPGNNNTTASGKNT-----PSDTRNDIQ--NDGIP-SKIKTDNENQOLK--KEFIS 1915
Db 1414 IEDSKRLNLEQKNTLKSYYKIKESVENDIKFAQEGISYYRYKAKKYKDDLESIKYKIR 1473
OY 1916 NMLQOPNDVPNDYSGNSSTVNTTTSRHNVDNNTTMSRDNMENMLLPSIHGML 1975
Db 1474 EEKEKFPSSPP---TTPSPAKTD-----EOKKEKFLPEL----- 1506
OY 1976 VSGEESTYNNVNSMNDIPINRDNNVYSGIDLINDISLGKRPIDITYEVLKRENELEFG 2035
Db 1507 -TNIEFLYV-NLNVKIDVLYN-----LAKKINDC-- 1534
OY 2036 TENVTRTSIQ--NVATTTNSDPIHNOLELFLHKMLDRHRDMCEKMKN---KEDLNLKIKEEM 2091
Db 1535 --NWEKDEAHVKITRLSDLKALIDKIDLEKFNHND--FDAIKRLINDDTKKMDLKGILLSTG 1590
OY 2092 NKENINNS-----GKTYNSDNKPSHNHY 2114
Db 1591 LVQNFPTIISKLIBGKFODMNLISOHOCV 1620

RESULT 9
MSP1_PLAFC ID MSP1_PLAFC STANDARD: PRT: 1726 AA.
AC P04934:
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MERZOITE SURFACE PROTEIN 1 PRECURSOR (MERZOITE SURFACE ANTIGENS)
DE (PMSA) (P195).
CN MSP-1.
OS Plasmodium falciparum (Isolate Camp / Malaysia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RX NCBI_Taxid=5835;
RX [1]
RP SEQUENCE OF 1-1103 FROM N.A.
RA MEDLINE=86205236; PubMed=3517809;
RA Weber J.L., Leininger W.M., Lyon J.A.;
RA "Variation in the gene encoding a major merozoite surface antigen of
RA the human malaria parasite Plasmodium falciparum.";
RL Nucleic Acids Res. 14:3311-3323(1986).
RN [2]

```

```

RP SEQUENCE OF 1104-1726 FROM N.A.
RX MEDLINE=88143999; PubMed=3278296;
RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
RA "Merozoite surface protein sequence from the Camp strain of the human
RA malaria parasite Plasmodium falciparum.";
RL Nucleic Acids Res. 16:1206-1206(1988).
CC -I- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -I- PRT: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage b/ and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL, X03831; CAN27446.1; -.
DR PIR, A23386; SAZOGM.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1
FT CHAIN 1
FT CAROXYD 20 1726 POTENTIAL.
FT CAROXYD 133 133 MERZOITE SURFACE PROTEIN 1.
FT CAROXYD 272 272 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CAROXYD 501 501 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CAROXYD 567 567 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CAROXYD 638 638 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CAROXYD 827 827 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CAROXYD 924 924 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CAROXYD 944 944 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CAROXYD 990 990 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CAROXYD 1016 1016 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CAROXYD 1114 1114 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CAROXYD 1221 1221 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CAROXYD 1613 1613 N-LINKED (GLCNAC . . .) (POTENTIAL).
SO SEQUENCE 1726 AA; 196197 MW; DD8AD45FA352BCF3 CRK 64;

Query Match 2.2%; Score 262.5; DB 1; Length 1726;
Best Local Similarity 19.0%; Pred. No. 0.00027;
Matches 372; Conservative 243; Mismatches 695; Indels 647; Gaps 90;

OY 444 YKLNKNDGVTDAFL---GLNNEKACKDITDGGKINKEVNSGGVYGGSGSGSGS 500
Db 25 YQELVKKLEALDEAVLTGYGLFHKEKMT-----LNDEITTKGA:AOGSTSGTSGS 76
OY 501 GTNDENKGFYRSEYQCPDCGVQHGKGNQWERKTKYKRMWSKLYIP:INKMVLKLS 560
Db 77 GTS-----GTSCTSGTSAQSGTSGS----- 97
OY 561 LKVVADMILKKNWEFCUJLONSSDQSGSVYTTGASG--GNSSEKE:YDEKCYKHNEY 618
Db 98 -----AQSCTSGTSAQSGTSGTSGTSPSS-----RSNTL 129
OY 619 QKVVNQGVEYEDDELKAGGGLCLIPNPKKNVEPSAKSQNNHADQCTFHDFPY---- 674
Db 130 PRSNTS-----SGA-----SPPADASDPAKS--YADLK HVRVRYLFTIKEL 169
OY 675 -----WVAHMLK--DS:H-----WRFKLSKSIDGKT 700
Db 170 KYPELFDLTNHLTLCDNTHGFKYLLDGYEETINELLKYKLNFFDLRLK:KLVNDYCAN----- 226
OY 701 MCRNCGCNKKDCFEKRVKQKET--EMKPIKDHFKTQEGIPGYYFTTLELILKQFLKE 758
Db 227 -YQQLPENLKIRANELDVLKLVFGYRKL- DNIDQNVGKMDYDK NKTITIANINELLE 284
OY 759 DTBENTENSLDDEAEELKHLQIKLENNENLAVVYNACTEDKTLMDKLNLHLELDATKC 818

```



RC TISSUE=Brain;  
 RX MEDLINE=93265134; PubMed=1303259;  
 RA Djibali M., Sellier L., Parry P., Bower M., Young B.D., Evans G.A.;  
 RT "A trichorax-like gene is interrupted by chromosome 11q23  
 translocations in acute leukaemias."  
 RL Nat. Genet. 2:113-118(1992).  
 [5]  
 RN SEQUENCE OF 1251-1538 FROM N.A.  
 RX MEDLINE=94215165; PubMed=8162575;  
 RA Gu Y., Alder H., Nakamura T., Schlichman S.A., Prasad R., Canaan O.,  
 Saito H., Croce C.M., Canaan E.;  
 RT "Sequence analysis of the breakpoint cluster region in the ALL-1 gene  
 involved in acute leukemia."  
 RL Cancer Res. 54:2326-2330(1994).  
 [6]  
 RN SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).  
 RX MEDLINE=95322025; PubMed=7598802;  
 RA Mbongkolle D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,  
 Rowley J.D., Diaz M.O.;  
 RT "The human MLL gene: nucleotide sequence, homology to the Drosophila  
 trx zinc-finger domain, and alternative splicing."  
 RL DNA Cell Biol. 14:475-483(1995).  
 [7]  
 RN SEQUENCE OF 1212-1603 FROM N.A.  
 RX MEDLINE=95315013; PubMed=7794749;  
 RA Marschalek R., Grell J., Lochner K., Nilson I., Slegler G.,  
 Zweckbronner I., Beck J.D., Fey G.H.;  
 RT "Molecular analysis of the chromosomal breakpoint and fusion  
 transcripts in the acute lymphoblastic SEM cell line with chromosomal  
 translocation t(4;11)."  
 RL Br. J. Haematol. 90:308-320(1995).  
 [8]  
 RN SEQUENCE OF 1421-1540 FROM N.A.  
 RX MEDLINE=94020842; PubMed=8414518;  
 RA Forster A., Rabblits T.H.;  
 RT "A method for identifying genes within yeast artificial chromosomes:  
 application to isolation of MLL fusion cDNAs from acute leukaemia  
 translocations."  
 RL Oncogene 8:3157-3160(1993).  
 [9]  
 RN CHROMOSOMAL TRANSLOCATION WITH GAS7.  
 RP MEDLINE=20183971; PubMed=10706619;  
 RX Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,  
 RA Wilson R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,  
 RA Williams T.M., Lange B.J., Felix C.A.;  
 RT "Detection of leukemia-associated MLL-GAS7 translocation early during  
 chemotherapy with DNA topoisomerase II inhibitors."  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).  
 CC -i- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.  
 CC -i- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -i- TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T AND B LYMPHOCYTES.  
 CC -i- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS  
 T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND MLLT1/ENL;  
 T(4;11)(Q21;Q23) THAT INVOLVES MLL AND MLLT2/AF4; T(9;11)(P22;Q23)  
 THAT INVOLVES MLL AND MLLT3/AF9; T(6;11)(Q27;Q23) THAT INVOLVES  
 MLL AND MLLT4/AF6; T(11;17)(Q23;Q21) THAT INVOLVES MLL AND  
 MLLT6/AF17; T(X;11)(Q13;Q23) THAT INVOLVES MLL AND MLLT7/AFX1;  
 T(10;11)(P12;Q23) THAT INVOLVES MLL AND MLLT10/AF10;  
 T(1;11)(Q21;Q23) THAT INVOLVES MLL AND AF10; T(11;19)(Q23;P13.3)  
 THAT INVOLVES MLL AND ENL; AND T(11;19)(Q23;P23) THAT INVOLVES MLL  
 AND GAS7.  
 CC -i- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION  
 FACTORS.  
 CC -i- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
 CC -i- SIMILARITY: CONTAINS 1 SET DOMAIN.  
 CC -i- SIMILARITY: CONTAINS 3 PHD ZINC-FINGER DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch).  
 CC -----  
 DR EMBL: L04284; AAA58669.1; -  
 DR EMBL: 269744; CAA93625.1; -  
 DR EMBL: 269745; CAA93625.1; JOINED.  
 DR EMBL: 269746; CAA93625.1; JOINED.  
 DR EMBL: 269747; CAA93625.1; JOINED.  
 DR EMBL: 269748; CAA93625.1; JOINED.  
 DR EMBL: 269749; CAA93625.1; JOINED.  
 DR EMBL: 269750; CAA93625.1; JOINED.  
 DR EMBL: 269751; CAA93625.1; JOINED.  
 DR EMBL: 269752; CAA93625.1; JOINED.  
 DR EMBL: 269753; CAA93625.1; JOINED.  
 DR EMBL: 269754; CAA93625.1; JOINED.  
 DR EMBL: 269755; CAA93625.1; JOINED.  
 DR EMBL: 269756; CAA93625.1; JOINED.  
 DR EMBL: 269757; CAA93625.1; JOINED.  
 DR EMBL: 269758; CAA93625.1; JOINED.  
 DR EMBL: 269759; CAA93625.1; JOINED.  
 DR EMBL: 269760; CAA93625.1; JOINED.  
 DR EMBL: 269761; CAA93625.1; JOINED.  
 DR EMBL: 269762; CAA93625.1; JOINED.  
 DR EMBL: 269763; CAA93625.1; JOINED.  
 DR EMBL: 269764; CAA93625.1; JOINED.  
 DR EMBL: 269765; CAA93625.1; JOINED.  
 DR EMBL: 269766; CAA93625.1; JOINED.  
 DR EMBL: 269767; CAA93625.1; JOINED.  
 DR EMBL: 269768; CAA93625.1; JOINED.  
 DR EMBL: 269769; CAA93625.1; JOINED.  
 DR EMBL: 269770; CAA93625.1; JOINED.  
 DR EMBL: 269772; CAA93625.1; JOINED.  
 DR EMBL: 269773; CAA93625.1; JOINED.  
 DR EMBL: 269774; CAA93625.1; JOINED.  
 DR EMBL: 269775; CAA93625.1; JOINED.  
 DR EMBL: 269776; CAA93625.1; JOINED.  
 DR EMBL: 269777; CAA93625.1; JOINED.  
 DR EMBL: 269778; CAA93625.1; JOINED.  
 DR EMBL: 269779; CAA93625.1; JOINED.  
 DR EMBL: 269780; CAA93625.1; JOINED.  
 DR EMBL: D14540; BAA03407.1; -  
 DR EMBL: L01986; AAA92511.1; -  
 DR EMBL: U04737; AAA18644.1; -  
 DR EMBL: S78570; AAB34770.1; -  
 DR EMBL: X83604; CAA58584.1; -  
 DR EMBL: S66432; AAB28545.1; -  
 DR EMBL: AF231998; AAC26332.2; ALT\_TERM.  
 DR TRANSFAC: T02337; -  
 DR MIM: 159555; -  
 DR InterPro: IPR001214; -  
 DR InterPro: IPR001487; -  
 DR InterPro: IPR001965; -  
 DR InterPro: IPR002857; -  
 DR Pfam: PF006028; PHD: 3.  
 DR Pfam: PF00856; SET: 1.  
 DR Pfam: PF02008; ZF-CXXC: 1.  
 DR PROSITE: PS50014; BROMODOMAIN\_2; 1.  
 DR PROSITE: PS50280; SET; 1.  
 KW Nucleic acid binding; Chromosomal translocation; DNA-binding; Bromodomain;  
 KW Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;  
 KW Alternative splicing.  
 FT DOMAIN 17 102  
 FT DNA\_BIND 169 180 ALA/GLY/SER-RICH.  
 FT DNA\_BIND 217 227 A.T HOOK (BY SIMILARITY).  
 FT DNA\_BIND 301 309 A.T HOOK (BY SIMILARITY).  
 FT ZN\_FING 1431 1482 A.T HOOK (BY SIMILARITY).  
 FT ZN\_FING 1484 1533 PHD-TYPE 1.  
 FT ZN\_FING 1566 1627 PHD-TYPE 2.  
 FT DOMAIN 1703 1748 PHD-TYPE 3.  
 FT DOMAIN 3840 3969 BROMODOMAIN (DIVERGENT).  
 FT DOMAIN 137 143 SET.  
 FT DOMAIN 561 564 POLY-GLY.  
 FT DOMAIN 568 571 POLY-PRO.  
 FT SITE 1444 1445 BREAKPOINT FOR TRANSLOCATION TO FORM MLL-

```
FT VARSPLIC 1407 1444 GAS7 ONCOGENE.
FT CONFLICT 144 144 MISSING (IN ISIFORM 14P-18B).
FT CONFLICT 317 379 E -> ELTQIIPCSMRKTHHDKTEPRFLAMSWCLN
GLINSELEKQAKRKREKGPPLTEKDYVRSPPRIK
VRIIPSSKRDTATIAKOLLOA -> VSSLINKSPKSG
KTRKHLQKRIKRLSDKALEISQGLFLQKQOPL
SNSYRGO (IN REF. 1).
R -> E (IN REF. 2).
Q -> G (IN REF. 5).
S -> SGTG (IN REF. 2 AND 7).
Q -> C (IN REF. 6).
Q -> H (IN REF. 4).
P -> S (IN REF. 4).
R -> G (IN REF. 2).
N -> D (IN REF. 2).
D -> G (IN REF. 2).
R -> A (IN REF. 2).
FT CONFLICT 556 556
FT CONFLICT 1487 1487
FT CONFLICT 1603 1603
FT CONFLICT 1616 1616
FT CONFLICT 1937 1937
FT CONFLICT 2181 2181
FT CONFLICT 3718 3718
FT CONFLICT 3759 3759
FT CONFLICT 3813 3813
FT CONFLICT 3901 3901
Query Match 2.28; Score 262.5; DB 1; Length 3969;
Best Local Similarity 18.28; Pred. No. 0.00074;
Matches 450; Conservative 292; Mismatches 888; Indels 837; Gaps 112:
OY 10 TODDAKHVDFEGOKVHD-----EVHGAKNVSELK-----GSLSLASTILG 52
DB 859 SKDRDADSVKDKSREDREREKENKREKREKKEKSSALYPVGRVSEKAYVG 918
OY 53 ETAFYVKSMTESKYTELEANSKRNPCKKDKGNDV-----DRFSYKE-----QAGYDN 102
DB 919 EDVATSSAK-----KATGRKSSHP--SGTDITVTLGDTTAVKTYILIKKGCN 968
OY 103 -KKKCSNGMTCAFP--RRLHLCKNFMNNSNDSSKAKHDLAEVMAKYESESITKH 160
DB 969 LKTNLDLGGTAPSLSEKKTLC-----LSTPSSSTVYKHSISGMLAQD----- 1014
OY 161 PKYOSKYGSPMPCTMLARSFADIGDIIIRGDLYLKKKKKKONGKETEREKLE---OKL 217
DB 1015 -----KLPMTKRVAASLKKAKAOLCKLEKSKSLKOTDQPAQ--QESDSSETSVRGPRI 1068
OY 218 KEIFKIHDLKDEAKQKRYNGEDPNFYKLRBDMWTANREVMGAMCSELDNSYFR 277
DB 1069 KHVCRRRAVALGRKA--VPPDDPITLSAL--PW--EEREKILSSMGND--DKSSTAG 1118
OY 278 A-----TCNDTGOGPSQTHNK-----CRCDDKKGANAG---KPK 308
DB 1119 SEDAEPIAPPIKPIKPYRNKAPQEPVKKGRSRSGCGCPGQVEDCGVCTNGLDKP 1178
OY 309 AGDGDVTVIYFYFVPOYLYLWMEBEMADFCRKKKKLENELEKQCRGKODETRYSRN 368
DB 1179 FGG-----RNIKKOCCMKRKCQNLOMMPK 1203
OY 369 GYDEQOTSRKGYVMGCTDFFACGSYENMIDONKQKODKKTYSKETSIDG----- 422
DB 1204 AYLOKQAKAVKKEKESK-----TSKKKSKESSVYKKNVDSQKPTP 1246
OY 423 GGRKRAVGGTTKYEGEKSEFYELKNDGYTVDAFLGLNNEKACKDITDGKINFEV 482
DB 1247 SARBDPAKKSSEPPRPKPYEKEKSEE--GNVSA-----PGPSKQATTPASRKSNOV 1298
OY 483 NSGGGVGGSGGSGTSGASGTDENKGTFRSEYICOPPCDCVOHKGNQMERKTKVKMR 542
DB 1299 SQPALYI--PPOPPTTGP--RKEVPKTPSEPKKQPP-----PESGPSQSKQKVAAPRS 1352
OY 543 WSKLYPFGKMWLLSLKAYVKDMMLKKNMKFECLONSSDSYGVYVTTGASGSE 602
DB 1353 IPVQKP-----KEKEKPPVAKQENAGTLLNTLT--LSNGSS 1389
OY 603 KKELYDEMCKYKHNVOKVNOGEVEEDDLKAGGLCIPN--PKKKREYSEAKSONH 661
DB 1390 KQKI-----PADGVIRIRIDPEKDECAENVWEMGGILITSPITPRVYCFICASSGH 1442
OY 662 ADIO-----KTFHDFEYVVAHMLKDSI--HWRTRKLSKCSIDSK-----TJMKCR 704
```

```
DB 1443 VEFYVQVCEPPEFHKFCLENERPLEDQLENMCCRCKFCYVCGROHQTQKLECNKR 1502
OY 705 NGCNKKC-----DCEKYNKQK-----EDEMK----- 726
DB 1503 NSYHECPGPNYPTPKTKKKKVVICTKVCCKSCGSTITPGKGMDAOHNSHDSLCIHDCAKL 1562
OY 727 -----PIKD-----HFKTOEGIPGEYFTLELILKLOFLK 757
DB 1563 FAKGNFCPLCKQCYDDDDDESKMMQCGCKDRMVHSCKNLSDEMEI-----LSNLP 1614
OY 758 EDTEBENTNSIDAEBAEELKLOKILKLENNNL--AYVNACTEOKTLMDKLNLHENDAT 816
DB 1615 ESNAYTVCNCEHREHRAERLALKEKLOISLKQVLTALLNSPTSHL-----RYROA 1667
OY 817 KCKOCPLEEDKSGRSRASPDIFFIPPEEKEDENEDDEDVVRDEETAKETTESGA 876
DB 1668 KPPDLNPTETESITRSRSPGPDPEVLTEVSKODQ-----OPLDEGKRMKMDGNY 1720
OY 877 TDTTTF--SLDVCPIVGKVLTKDNESLQDACSILKYGNNRSLGMRCVTPSGEPPTSSDKNGA 935
DB 1721 TSVLEFSDDIYKILQAAINSD-----GGQ-----PEIKKANGMVKSFF 1758
OY 936 ICVPPRRRLYIKKIYDMATTESPQASGSEASSTSGSTTPPDSKELLKAFVESALET 995
DB 1759 I-----ROMERYFPWFYSKSRFEMEPKNKYSSNG-----MLPNVALPPSDHN 1801
OY 996 FELMHRKKE-----KAAV--AOEGAGHLP-----RVEGSPREY 1028
DB 1802 YAOQDERRENSHTQOPPLMKKIIPAKPKGPEBDSPTPLPPTPILSTDRSDEDSPEL 1861
OY 1029 DPEDLKKEKIPDGLRQMFYTL-----GDYRDIIFSGSNDTTSV-----SKDTPSSS 1076
DB 1862 NPPGIEDN-----RQCALCLFYGDSDSANDAGRILYIGQENMTHVACALMSAEVEFD 1914
OY 1077 NDNKNIYLLASGSTEDEBEMANK-----YKEIKNF-----RK--C 1110
DB 1915 DGSLEKNHMAVIRGKOLCEFCQKPGATVGGCLTSGTSNYHFMCSRAKNCFVLDCKKYVC 1974
OY 1111 STERSAPNLVSPOTMMENNKYIMHGVN-----ALTSKDKLAKGVKKPKQIENPENL-- 1165
DB 1975 QRHDLKIGEYVPE-----NGFEVFRVYFVDFEGISLRKRLNGLF-----PENIMH 2021
OY 1166 -----WDEANKRPKPOYQYTNVKLDENSG-----TSP 1193
DB 2022 MIGSMITIDCLGILNDLSDCEDKLPPIGYQSRVYVSTTDARKRCVYICKIYECRPVYEP 2081
OY 1194 RTTQTOASSDNTPPTTLHFVNRPTYPFRFEMGSGFCREKRRKLQOT--KVDCQVENG 1250
DB 2082 DINSTVEHDEN--RTIAH--SPTSFT-----BSSKESQNTAEIISPPSPDRPHSQT 2130
OY 1251 VGRCSGDG--EACDSISTHDYSTVSPFNCPC----- 1280
DB 2131 SGSCYIHYISKVPRIRTPSYS--PTORSPPGCRPLPSAGSPPTTHETIYVGDPLLSGLR 2188
OY 1281 -----GKHCSYRKWIERRKIEF--HKQSNAYGOQKTDATRNNGNTDEKCEKTELETPD 1333
DB 2189 SIGSRHSTSLSS--PQSKRLRIMPRTGNTYSRNNAVSYTGTADBLE-----S 2238
OY 1334 AAKFLERLKNPCKTKNEKYGDDIDFEKDSKTPQHTKCGCPKFKNCOQNGVSGLN 1393
DB 2239 SAKVVDHVL--GPLNLSSTSLG-----ONT-----STSSMLQTVTVYVANKN 2277
OY 1394 GNCGDGKSIDAKEIKAKMSSSTVDVVMRVSDMDVTMFEQDLK-----DACQAHANFKG 1446
DB 2278 SHLDGSSSSSEMKQ-----SSASDLVSK-----SSSLGKTKYLLSSSSSESAHNVAYPG 2327
OY 1447 IRK-----DVMKGYVGVADICEOTININERTDGEKTYQIRALFKRVVENFLE 1493
DB 2328 IPKLAPQVHNHTSRELNVSKG-----SFAEPSSVS--FSSKELASPHLHLRGQRNDRD 2380
OY 1494 DY-----NKINDK-----ISHCIKGBGSKCINGCKENSK--CLEKNIK 1531
DB 2381 QHTDSTQSANSSPDEDEVEVTKLISGMSNRSSIINEHMGSSSRDRQKGGKSKETEK 2440
```

```

OY 1532 -----KIAEMENIKRRFNDQYENKQDPQYVNAKSLLEELIPKIAVYVNDQDNVIX 1579
DB 2441 HSKSFLREGVQYTTGEEGNLKEPFMD-----EVLPF----- 2472
OY 1580 LCVFENSKGCTLTSTONNKNENDADCMKLGLYAKNCPGKPSGKOSDKCEPPLPDE 1639
DB 2473 ---YMGQRCC-----NNVSSD-----KIGDKGLSMGPV-----KAPPMQVE 2506
OY 1640 ---EDONPEENTLPPKFCPTTPPREKGGCTGCKNEKKEKDEKESSEPAKESGPA 1695
DB 2507 GSAKELQAPRRKRTVK-----VTLTP-----LKMENESQSKNAKLESS--- 2543
OY 1696 AEPAPAESEETENFP-----EPGTPAPAPSPAPPTPT----- 1734
DB 2544 ---PASPIQIESTSPTEPLASASENPGDGPVAPSPNNNTSCDSQSNNYONLPEVDRLNML 2600
OY 1735 PEPAPQADEPDPSTLIQTTIPFGVALAGSIAFLFKKTKTASVGNLFOILQI---PK 1790
DB 2601 PGKPRQED-----GSFKRRYPRSAARSMFGLPLPGVRSY 2640
OY 1791 SDYDIPLKSNRYIPYSDRYKGTYYMEGD-----SDCKYAFMSDTTDTTSSES 1843
DB 2641 GREDIPFYSSS-----TGKRRKRSAGQYDAGADLSTDEDLYYNFRVYISGG 2693
OY 1844 EVELDINDIYVPS-----PKYKTLLEVLPERSGNNTTASGKNTPSDTRNDIQNGIPSS 1899
DB 2694 E-ERLASHNLFREEOCQDLPKISOLDGV-----DGTESDTSVYATTKRSSQ---IP-- 2741
OY 1900 KITDENWMLKKEFTSNMLQONPNDVNTSGNSTNTNITTSRHVNDN---NTNTT 1955
DB 2742 -----KNGKNGENGENKIDRPEDAGEKEHYTKSSVGHKPKKNDCHSRYKVT 2791
OY 1936 MSRDNMEENL-LTPS---IH-----DGNLYSGEEXSYNNVAVNSMNDIPINDNNVYSGI 2006
DB 2792 QOODSLEADLSSLESSRRYHTSTPSDKNLD---TYNTELLKSDS-----NNN----- 2837
OY 2007 DLINSLSGRPIIDYDEVLKKRENDLFGTEYTKRTSTONVAKTNSDPIHQLFPHK 2066
DB 2838 ---SDCCGNILPSDIMDFVLK-----NTPSMQALGESPESSSELMLNGEGLIGLD 2884
OY 2067 LDRHRDM 2073
DB 2885 SNREKDM 2891

RESULT 11
MSPL_PLAUF
ID MSPL_PLAUF STANDARD: PRT: 1701 AA.
AC P13819.
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MERZOITE SURFACE PROTEIN 1 PRECURSOR (MERZOITE SURFACE ANTIGENS)
DE (PMSA).
GN MSP-1.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86142939; PubMed=2449612;
RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
RA Brown G.V., Anders R.F., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens
RT of Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 27:291-302(1988).
CC -I- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -I- PUT: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR ANTIGENS OF
CC MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL: M19143; AAA29653.1; -.
DR PIR: A54498; A54498.
DR InterPro: IPR000561; -.
DR Pfam: PF00008; EGF; 1.
KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1
FT CHAIN 20 1701
FT CARBOHYD 110 110 (POTENTIAL).
FT CARBOHYD 239 239 (POTENTIAL).
FT CARBOHYD 470 470 (POTENTIAL).
FT CARBOHYD 536 536 (POTENTIAL).
FT CARBOHYD 607 607 (POTENTIAL).
FT CARBOHYD 802 802 (POTENTIAL).
FT CARBOHYD 899 899 (POTENTIAL).
FT CARBOHYD 919 919 (POTENTIAL).
FT CARBOHYD 965 965 (POTENTIAL).
FT CARBOHYD 991 991 (POTENTIAL).
FT CARBOHYD 1089 1089 (POTENTIAL).
FT CARBOHYD 1196 1196 (POTENTIAL).
FT CARBOHYD 1588 1588 (POTENTIAL).
SQ SEQUENCE 1701 AA; 193719 MW; 3920875E73038552 CRC34;

Query Match 2.1%; Score 258; DB 1; Length 1701;
Best Local Similarity 19.0%; Pred. No. 0.0041;
Matches 357; Conservative 241; Mismatches 660; Indels 620; Gaps 90;

OY 565 KDMMLTKMKWKEFLQNSDGSVGSVYVTG-----ASGNSKRLLYDEMKCYKRN 616
DB 48 KKMVNLNEGSTGTAVY-TETPGSSGVSITSGSVASVYASVAGSG- 91
OY 617 EYQKVVQGEVEDEDELGAGLCLLPNPKKKEVSEAKSNHNDIKOTFHFPEYV- 674
DB 92 -----GSV-----AGSGSNGSRTPNDSSDSNTKT---YADLKHRYQNTYFTIK 134
OY 675 -----WVAHMLKDSIHWRTKRLKSCIS-----IGTKMKCRNGC-N 708
DB 135 ELKYPELFDLTJNML--TLSKNVDFGKYLDIGYEELNELLYKLFYLLLRKALDANACN 192
OY 709 KKQDC-FEKWVQKETE-----WKPIKHFKTOBEPICGYFTTILELIKLOFLKE 758
DB 193 SYQIIPFNLIKIRANELDYAKTIVFGYRKL-DNIKQNVKMDYIKK KTTIANINELLE 251
OY 759 DNEENTENSLDAEAEELKLOKILKLENNNLAAYVNACTEOKTLMDFLLNHELNDATGC 818
DB 252 GSKKTIQDKKNKNDNEGKKLL-----YQAYNLFYTNKQLOEHNLLIVLEKRIDPLKKN 306
OY 819 KDCP-LPEEDKSRGRSADPS---PDIPPREPEKEDDENEDDEIVRDEETAKETT 872
DB 307 ENIKKLEDIDKIKTDAENPTGSKPN---PLPENK-----KIVGEHKKIKE-- 352
OY 873 EGSADTTTSLDYCPYGVVLTKNDSL-QDACSLEY--GGNSRLGIRCYPPSGEPTTS 929
DB 353 -----IACTIKFNISLPTDPLELEYLRKKNKYD---VTPKSODPKR 393
OY 930 SDK-----NGAICVPPRRRLRIKIKYDWTAKTESPOASGEASSTSGSTTPDSKCA 982
DB 394 SVOIKRVPIPNKIVPLP-----LTDHNSLAANDKNS-YGDLNPDPTK 439
OY 983 LKAAYESALETFFL-----WHRYKEEKVAQEGAGHILPRVEESPEYD 1029
DB 440 INEKITONKREKIFINNIKKQIDLEENKINHTKQNKLLD-----YKSKQYE 491
OY 1030 P-EDLKEGKIDGFLRQMFYTLGDIROLIFSGSNDTTSVSK-----DIPSSNDNLKNI- 1083

```



```

Db 492 ELLEKFEYMKNNNDK-----DYYDKIFS-ARYTYNVEKORYNNKFSNNSSNNVYVQ 543
Oy 1084 -----VLLASSTEOEREMKMYKELKRNKRCSTERSAPNLVSHQTMWENNGK 1132
Db 544 KLKALSLLEYSLRKIGISED---FNHYTLKLGLEADIKKLEIKSEKNIKLEKFK 600
Oy 1133 YIWMGWCALTSKDKIAKVER-----KPKIE-----NPNLMDANKK 1172
Db 601 GLTHSANSLSLEVSIDIVKLOVQKVLKIKKIEDLRKLEFLKNAQLKDSJHVNIYIPQK- 659
Oy 1173 PKPPOYQYTNV-----KLDE----- 1187
Db 660 ---PEPYLLYLKKEVDLKEFIPKVKMLKEQAVLSITQPLVAASETTEDEGSHSTHT 716
Oy 1188 ---NSGSPRTQTOQASSNPTJTLTHFYKRPYFPMFEMGESFCREKKRLKOIKYVCK 1245
Db 717 LSGGETEVTETEETETETETVHTTVTITLPP-----KEESAPEKVKV--- 759
Oy 1246 VENGDVGRCSGDGA-----CDSISTHDYSTVPFNCPGGKHCSSYRKWIERKKI 1296
Db 760 VENSIEHKSNNSQLTKTYVYLKKLDEFLTKSYI-----CHKYILVNSSMDOKLL 810
Oy 1297 E-----FKQSN-----AYGQKTDATRNKNTFDKE---FCK 1326
Db 811 EYVNLTPREENELKSCDPLDLFNIONNIPAMYSLYDSMNIDLQHLFELYQKEMITYLH 870
Oy 1327 TLETWPDAAKFLERLK---NGSPCKTNKEVGGDDIDEPKSKTFPHTEVYCGPCPKFTNCON 1384
Db 871 KLKEBNHKLLEBKOKQITGTSSTSSP-GNTTVN-----TAQSAVH-----SNSON 915
Oy 1385 GNCVSGSINGNCDDKSIDAKEIAKMRSSITDVVARYSDNDNTFEEDDLKDACOHANIF 1444
Db 916 QOSNAS-----STNQNGVAVSSGPVAVESHDPPLTVLSTISNDL 954
Oy 1445 KGIRKDWKCGYVCGVDCEPTNINERTDGEKEYIQIRAL-FKRWVENFLEDYK-INDKI 1502
Db 955 KG-----IVSLNLGNKTKVNPPLISTTEKEKEVENLKKNDDYFNDI 999
Oy 1503 SHCJIKGSGKICNG---CEKNS-----KCLEMKTEKIAEMENKRRFNQYENK---D 1551
Db 1000 KOYFK---SNSAVITGLTQKALNDELKIKLKDITQLSFDLYNKTKLKLDRLFNKKKELG 1057
Oy 1552 QPDYVNASITL---ELLIPKIAVVDQDNVIR-LCVFENSKCCTLISNTONKKNDAIDCM 1607
Db 1058 QDKQIKKLTLLKEOLESEKLSLNNPHNVLQNFVSFEFKKKEALAELENTLENTKI--L 1115
Oy 1608 LKRLGVAKKNCPRKPSGKQSDCKEPPLPDEEDONPEENTLEPPKFCPTTQPEEKGG 1667
Db 1116 LKHVKGVLKYYNGESS-----PLKTLSEVSIQTEDNANLEKF-----RALSKIDG 1161
Oy 1668 ETCGNKE--EKK-----DEKKESEPPAKESGPALEBPATTAESEGETNF 1713
Db 1162 KLANDLHGKKKLSFLSSGLHLLITELKEVTKNNKNTGNSSENKKNKYNALKESTENFLP 1221
Oy 1714 EPPGTGPAAPSTPAPPTPD-TPPPLRPOADEPFDSTILQTTIFGVALAGSIAFLFLK 1772
Db 1222 E-----AKVTYVTPPQPDVTPSPLSVYSGSSGSTKEEQIOP-----TSGSL----- 1264
Oy 1773 KKTIVASGNLFQIIQIIPKSDYDITPLKSSNRYIPYVSDRYGKTYIWEKGS-DEKX- 1829
Db 1265 -----LTLEQVVQVQLONYDEEDDSL-----VVLPIFGSESDNDEYLD 1301
Oy 1830 -----AFMSDITD--VYSSESEYE-----ELDINDIYYPG 1857
Db 1302 QVWTEGAISVYMDNITLSGFENEYDVIYKPLAGVYRSILKQIEKNITTFENLMDINSR 1361
Oy 1858 SPKYKTLIEV-----VLEPSGNNTASGKNT-----PSDTRNDIO- 1892
Db 1362 LKRRKYFLDVLESDLMQFKHISMEYIIEBSFKLNSQKTKTLKSYKIKESVENDIKF 1421
Oy 1893 -NDGIP--SSKITDNEWQOLK--KEFISNMLQNOPNDVPNDYTSNGSSNTNTITTSRRHV 1948

```

```

Db 1422 AOEGISYEVKYLAKYKDDLESIKKIVKEKEKFPSSPP---TTPPSPAKTD----- 1469
Oy 1949 DNNTWTMSRONMEENLLPSIHONLTSGEYSYVNMVNSMNDIPN-----RDNNV- 2002
Db 1470 -----EOKRESKLPPL-----TNIETLYN-NUVKIDYILINLAKINDCAVE 1512
Oy 2003 ---YSGIDLINDLSGGKPIDYD-----EVLKRENELEFGTENTKRTS-----T 2044
Db 1513 KDEAHVKITKLSLDLKAIDKIDLFPKNTNDPFAIKKLIN-----DDTKMDLGLKLLSTGLV 1567
Oy 2045 QNVAKTNSDPDIIHOLEFHFKWIDRHMRDCEKMKRKE-----ILNKIKE 2089
Db 1568 QNFPNTIISKLEEG---FQDMLNISOHQCVKQKQCPENSGCFRHLDERECKCLLN-YKQ 1623
Oy 2090 EWNKENINSGKTYNSDN 2107
Db 1624 EGDK-CVENPMPCTNENN 1640

RESULT 12
ID ATRX_MOUSE STANDARD: PRT: 2476 AA.
AC 061687;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR ATRX (X-LINKED NUCLEAR PROTEIN)
DE (HETEROCHROMATIN PROTEIN 2) (HPI ALPHA-INTERACTING PROTEIN) (HPI-BP38
DE PROTEIN).
GN ATRX OR XNP OR HPIBP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98213653; PubMed=9545503;
RA Picketts D.J., Tastian A.O., Higgs D.R., Gibbons R.J.;
RT "Comparison of the human and murine Atrx gene identifies highly
RT conserved, functionally important domains.";
RL Mamm. Genome 9:400-403(1998).
RN [2]
RP SEQUENCE OF 325-1176 FROM N.A.
RX MEDLINE=97133299; PubMed=9978696;
RA le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,
RA Jeanmougin F., Losson R., Chambon P.;
RT "A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic
RT control of transcription by nuclear receptors.";
RL EMBO J. 15:6701-6715(1996).
CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PHD-FINGER DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF026032, AAC08741.1; -.
DR EMBL, X59643; CA67962.1; -.
DR MGD; MGI:103067; Xnp.
DR InterPro; IPR000330; -.
DR InterPro; IPR001650; -.
DR Pfam; PF00176; SNF2_N.1.
DR Pfam; PF00271; Helicase_C.1.
DR DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding.
FT DOMAIN 219 267 PHD-FINGER.
FT NP_BIND 1579 1586 ATP (POTENTIAL).

```





QY	1777	ASVGN-----	-----LQGIQIPKSDYDITFLKS--NRRIPYSDPKRKKT	1816
Db	1569	KSPESGICLHCHGGLGKTTLDVVSFLHTVLCDLSTALVCPINTALAMNNEFK	-----	16255
QY	1817	YIVMEGSDDEKCAFAMSDTDTVYSESEY---EELDINDIYVPGSPKKTILEVLEPSG	18737	16255
Db	1626	--WQGLINDMEKLEVSLELAVKKRPGRSYWLQRWQDGGVMILGYEMKRNL-----	1674	1674
QY	1874	NNITASCKNTPSDTRNDIQNDGI---PSSKTTDNEMNOLKKEFTISMLQNDPNDPNDY	19299	19299
Db	1675	---AQRNKKSRKLKDFIKFKAALVDGPPFVVCDE-----GHILKNEASAVSKAM	1720	1720
QY	1930	TSGNSSTNTNTTTSRNVNNTNTMSDNDNEENLPSHDGNLYSGEESYVNVNV-	19888	19888
Db	1721	NSISRRRIITGTP-----LNNLIT-----ETHCNVPRIK	1751	1751
QY	1989	-NSMNDIPINRDNVNVSGIDLINDSLSGKPIDIDYDEVLRKNEELF---GTENTKRTS	2043	2043
Db	1752	ENLGSIIKEFRNREIN--PIQGQCADSTWVDV--RVKKKRAHILYEMLAGCVQRK---	1803	1803
QY	2044	TQNVAKTTNSDPIN-----QLELFHKLLDR-----	2069	2069
Db	1804	-DYTALTKFLPPEKHEVYLAVRMTALQCKLYOYLHLTGVSNGSTEGRGKAGAKLFDFQ	1862	1862
QY	2070	--HRDMCEKMKNKEDILNLTKEEMKNENLNSGKTVSDNKRSHNHVINTDVSIOIDMDN	2127	2127
Db	1663	MLSKWTHHPMLCLDITDYSKNKGYFEDSDMEFIASDSD-----ETSKLSLSD-EK	1912	1912
QY	2128	PKRTNEITNMDTQNDKSTMTQITLDDLEKYNDPYDYEDDIYHDVYEK-----	2178	2178
Db	1913	KKPRGKKGKKDSSSGSGSD-----NDVEIKVMNRSRSG	1948	1948
QY	2179	--SSMDIYVDHNNTYSNNNDVPTKM	2202	2202
Db	1949	GDGNMD-----TGNNPSVSLKL	1966	1966

RESULT 13

TPR\_HUMAN STANDARD: PRT: 2349 AA.

AC P12270;

DT 01-OCR-1989 (Rel. 12, Created)

DT 01-OCR-1996 (Rel. 34, Last sequence update)

DT 01-OCR-2000 (Rel. 40, Last annotation update)

DE NUCLEOPROTEIN TPR.

GN Homo sapiens (Human).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI\_TaxId:9606;

OX [1]

RP SEQUENCE FROM N.A.

RP MEDLINE=93064711; PubMed=1437155;

RA Mitchell P. J., Cooper C. S.;

RT "The human tpr gene encodes a protein of 2094 amino acids that has extensive coiled-coil regions and an acidic C-terminal domain.";

RL Oncogene 7:2329-2333(1992).

RN [2]

RN REVISIONS, AND CHARACTERIZATION.

RP MEDLINE=95096166; PubMed=7798308;

RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T., Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Geran L.;

RT "tpr, a large coiled coil protein whose amino terminus is involved in activation of oncogenic kinases, is localized to the cytoplasmic surface of the nuclear pore complex.";

RT J. Cell Biol. 127:1515-1526(1994).

RN [3]

RP SEQUENCE OF 1-142 FROM N.A.

RP MEDLINE=88262257; PubMed=3387099;

RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;

RT "tpr homologues activate met and raf.";

RL Oncogene 2:617-619(1988).

RL

```
CC -1- FUNCTION: COMPONENT OF THE CYTOPLASMIC FILIBLS OF THE NUCLEAR PORE
CC COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS
CC IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
CC TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
CC COMPONENTS, INCLUDING P62.
CC -1- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
CC BRAIN. LOWER LEVELS IN HEART, LIVER, AND KIDNEY.
CC -1- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
CC OR RAF GENES.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X66397; CAA47021.1; -.
DR EMBL; Y00672; CAA68681.1; -.
DR PIR; S00928; S00928.
DR MIM; 189940; -.
KM Hepad repeat pattern; Colled coll; Proto-oncogene;
KM Chromosomal translocation; Nuclear protein; Transort.
FT DOMAIN 78 360 COILED COIL (POTENTIAL).
FT 422 571 COILED COIL (POTENTIAL).
FT DOMAIN 575 628 COILED COIL (POTENTIAL).
FT 758 805 COILED COIL (POTENTIAL).
FT DOMAIN 834 869 COILED COIL (POTENTIAL).
FT 934 979 COILED COIL (POTENTIAL).
FT DOMAIN 1004 1064 COILED COIL (POTENTIAL).
FT 1138 1166 COILED COIL (POTENTIAL).
FT DOMAIN 1196 1241 COILED COIL (POTENTIAL).
FT 1262 1304 COILED COIL (POTENTIAL).
FT DOMAIN 1354 1434 COILED COIL (POTENTIAL).
FT 1476 1595 COILED COIL (POTENTIAL).
FT DOMAIN 527 530 POLY-SER.
FT 1833 1836 POLY-GLU.
FT DOMAIN 1957 1964 POLY-ASP.
FT 2295 2298 POLY-SER.
SQ SEQUENCE 2349 AA; 265600 MW; AFD6885CEDCA9EF CRC64;

Query Match 2.1%; Score 249.5; DB 1; Length 2349;
Match Local Similarity 16.8%; Pred. No. 0.0015;
Matches 390; Conservativa 340; Mismatches 771; Indels 815; Gaps 97;

OY 38 VELGSLSLA---SILEETATPVKSMQTESKYTELIEANSKRNPCKDKGNDVDFSV 94
DB 92 LTEKKKELEIADRRIALAIOSQTPTRKEELEAKRDLRIINELSO-ELFYLETDYKRL-- 148
OY 95 KEAGYDKKMKMSNGMTCAFFRRLHLNKNKFNPNMNSNDSSKAKHDLLAECYMAAKYEGE 154
DB 149 -----NEKLKESN-----TTTGELQDLKDELQASDVSKYREK 181
OY 155 STKTPYPRYDSYPPSSDPMPCTMLARSA---DIDITIRGRDLYGNKK-----KKQN 204
DB 182 RDEQEKELLSQNTMLANTELKKTDTDELALGKREKNELTELKCNLENKKEVSRLEEDQN 241
OY 205 GKEETEREKLEQKLEIFPKKHIDNLKDEKA--OKRYNGDED-----PNYYKLREDWWTANR 257
DB 242 GUKTISENHLQKAYEVDLTKLKE-AKEQDASMEKFNHNLMAHIIKISNLTK----- 290
OY 258 ETWVGAMTCKSELNDSYFRATCNDTGGQPSQTHNKKRCQDKDGAMAGKPRAGDGVYIV 317
DB 291 -----SAADS-----EAKSNELTRAVELHKLKL-----EAGEANKAI- 324
OY 318 PTYPDYVQYLRFMEFNAEDCRKKKKKLEMLELEKQCRKDKSDERYTCSRMGYDCEQITIS 377
DB 325 -----ODHILEVQSSDKDMEKEMLEKIGRLEKL-----EVANDLLSATK 364
```



```

RX MEDLINE-92357776; Pubmed-1496004;
RA Adams J.H., Sim B.R., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.;
RL "A family of erythrocyte binding proteins of malaria parasites.";
CC Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
CC -FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
CC DETERMINANT.
CC -SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -SIMILARITY: HIGH, TO P.VIVAX DUFFY RECEPTOR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M90694; AAA29603.1;
KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane;
KM Multigene family.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1153 DUFFY RECEPTOR, BETA FORM.
FT DOMAIN 22 1085 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1086 1106 POTENTIAL.
FT DOMAIN 1107 1153 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 348 348 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 626 626 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 847 847 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 856 856 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 900 900 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 910 910 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 935 935 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1153 AA; 130471 MW; 6497BD16CE7BEC01 CRC64;

Query Match 2.1%; Score 249; DB 1; Length 1153;
Best Local Similarity 19.0%; Pred. No. 0.00064;
Matches 197; Conservative 125; Mismatches 343; Indels 370; Gaps 45;

```

```

Db 387 YRWIREMGDMYSELPEQRKIKKCDRLKLYTNLRICITMS-----P 428
Oy 388 CTDFEACGSEYNNIDNOROFKQKTKETISDGGKKRAVGTTKYE-----GYE 440
Db 429 CND---SCKLYDQMTTKKKKKQWVLS--TKSSYKKGQIITENTITAYDILKQELNGFN 483
Oy 441 KSFYE-----KLK-----NDGY 452
Db 484 EWFENEINKRDYNYIDICLCADEPNKNTQEHKLKLSAPKLETORSHSTIQPMSSGA 543
Oy 453 GTVAFLAGLLN-NEKACKDITDGGKINFEVNSGGVGGVGGSGTSGASGT-NDENKGT 510
Db 544 EKVGDGLAHGNIINDAAVKSTTD-----EAKKDGGONGMTVAESNIKGDNIENEA 596
Oy 511 YRSEYOCPCPCPGVGHGNGNWEKTKYKKRWKSLKYPKNGWLLKSL-----KYVKD 566
Db 597 NVDTY-KVTVRSADTKGA-----TDITTEGEKLTNTSYSGSEITVYKINIPDGIYKD 649
Oy 567 MMI-----LKKRWKEFCLTONSSDGSVGSVYTTGASGNSKKELYDEMKCYKHN-- 616
Db 650.VSAAVENSENPLETKHKIFEPKSDNSDSENS-----GSM-----EFKATSSNPI 694
Oy 617 --EVQKVVQGEVEDEDELKAGAGLCILPNPKKKEVSEAKSONNHADIDKTHDEFFY 674
Db 695 TEAVESSSAEGQVQEDSAHRSVNTG-----RDNSTISAATSD----- 731
Oy 675 WVAHMLKDSIMWRTKRLKSCISDCKTMKCRNGCNKKCCCFKWKOKETEMKPI----- 728
Db 732 -----DGLSSGDKRVESSLT-----IENADGGDPYQGSILMN 763
Oy 729 -----KDFKTOES-----IPGGYPTTLELILKLOFLKEDTEN 763
Db 764 LNDPSVAGGSGKSHIKTEENEGSOAELDKGNVQIAEGRATITREYQPERPDLSDTONGN 823
Oy 764 TENSIDAE-----AELKHLQKLIK-----LENENNLAAYNAGTEQXTLMDKL 807
Db 824 VPRGNNKNEGATATLGSLESNESYHKTIDNTTHGLENK-----NGNEKPEFORHDF 877
Oy 808 LNHE-LNDATCKCKCPLEEDKSGRSADSPDFFIRPEKEDEDEND--DEDEVARD 864
Db 878 MNNDMLNDQTS-SPQTSDDTSSNQTSDDQTS-----NOTSSDQTSDDQTSDDQTS 930
Oy 865 EETAKETTESATDT 879
Db 931 QTSNQNQSSDQTIPT 945

RESULT 15
ATTRX_HUMAN STANDARD; PRT; 2375 AA.
ID ATTRX_HUMAN P46100; P51068; O15886;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR ATTRX (X-LINKED HELICASE II) (X-LINKED
DE NUCLEAR PROTEIN) (XNP).
GN ATTRX OR RAD54L OR XR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97123494; Pubmed-8968741;
RA Pickrelts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.,
RA Gibbons R.J.;
RT "ATTRX encodes a novel member of the SNF2 family of proteins: mutations
RL point to a common mechanism underlying the Attr-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-97386582; Pubmed-9244431;
RA Villard L., Losal A.M., Cardoso C., Proud V., Chitroni P.,

```

RA Coliaux L., Schwartz C., Fontes M.;  
RT "Determination of the genomic structure of the XNP/ATRX gene encoding  
RT a potential zinc finger helicase."; Genomics 43:149-155(1997).  
RL [3]  
RN SEQUENCE OF 743-2375 FROM N.A.  
RX MEDLINE-9517911; PubMed-7874112;  
RA Stayton C.L., Dabovic B., Gulsano M., Gecz J., Broccoli V.,  
RA Giovannazzi S., Bossolasco M., Monaco L., Kastan S., Boncinelli E.,  
RA Bianchi M.E., Consalez G.G.;  
RT "Cloning and characterization of a new human Xq13 gene, encoding a  
RT putative helicase."; Hum. Mol. Genet. 3:1957-1964(1994).  
RL [4]  
RN PRELIMINARY PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE-94214473; PubMed-8162050;  
RA Gecz J., Pollard H., Consalez G., Villard L., Stayton C.,  
RA Millaesau P., Khrestchatsky M., Fontes M.;  
RT "Cloning and expression of the murine homologue of a putative human  
RT X-linked nuclear protein gene closely linked to PKX1 in Xq13.3."; Hum. Mol. Genet. 3:39-44(1994).  
RL [5]  
RN SEQUENCE OF 2284-2375 FROM N.A., AND VARIANTS ATR-X.  
RX MEDLINE-95211835; PubMed-7697714;  
RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;  
RT "Mutations in a putative global transcriptional regulator cause X-  
RT linked mental retardation with alpha-thalassemia (ATR-X syndrome)."; Cell 80:837-845(1995).  
RL [6]  
RN VARIANTS ATR-X.  
RX MEDLINE-97467722; PubMed-9326931;  
RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Aisenbauer B.,  
RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,  
RA Levin M.L., Masuno M., Nerl G., Pierpont M.E., Slaney S.F.,  
RA Higgs D.R.;  
RT "Mutations in transcriptional regulator ATRX establish the functional  
RT significance of a PHD-like domain."; Nat. Genet. 17:146-148(1997).  
RL [7]  
RN VARIANT JM GLN-2014.  
RX MEDLINE-96224392; PubMed-8630485;  
RA Villard L., Gecz J., Mattel J.-F., Fontes M., Saugier-Verber P.,  
RA Munnich A., Lyonnet S.;  
RT "XNP mutation in a large family with Juberger-Marsidi syndrome."; Nat. Genet. 12:359-360(1996).  
RL [8]  
RN -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES  
CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN  
CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-  
CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X  
CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE  
CC PSYCHOMOTOR RETARDATION, CHARACTERISTIC FACIAL FEATURES, GENITAL  
CC ABNORMALITIES, AND ALPHA-THALASSEMIA.  
CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERGER-MARSIDI SYNDROME  
CC (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY  
CC SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,  
CC MICROGENITALISM AND EARLY DEATH.  
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 PHD-FINGER DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL, U72938; AAB40699.1; -;  
DR EMBL, U72935; AAB40699.1; -;  
DR EMBL, U72904; AAB40699.1; JOINED.  
DR EMBL, U72907; AAB40699.1; JOINED.

DR EMBL, U72908; AAB40699.1; JOINED.  
DR EMBL, U72909; AAB40699.1; JOINED.  
DR EMBL, U72910; AAB40699.1; JOINED.  
DR EMBL, U72911; AAB40699.1; JOINED.  
DR EMBL, U72912; AAB40699.1; JOINED.  
DR EMBL, U72913; AAB40699.1; JOINED.  
DR EMBL, U72914; AAB40699.1; JOINED.  
DR EMBL, U72915; AAB40699.1; JOINED.  
DR EMBL, U72916; AAB40699.1; JOINED.  
DR EMBL, U72918; AAB40699.1; JOINED.  
DR EMBL, U72919; AAB40699.1; JOINED.  
DR EMBL, U72920; AAB40699.1; JOINED.  
DR EMBL, U72921; AAB40699.1; JOINED.  
DR EMBL, U72922; AAB40699.1; JOINED.  
DR EMBL, U72923; AAB40699.1; JOINED.  
DR EMBL, U72924; AAB40699.1; JOINED.  
DR EMBL, U72925; AAB40699.1; JOINED.  
DR EMBL, U72926; AAB40699.1; JOINED.  
DR EMBL, U72927; AAB40699.1; JOINED.  
DR EMBL, U72928; AAB40699.1; JOINED.  
DR EMBL, U72929; AAB40699.1; JOINED.  
DR EMBL, U72930; AAB40699.1; JOINED.  
DR EMBL, U72931; AAB40699.1; JOINED.  
DR EMBL, U72932; AAB40699.1; JOINED.  
DR EMBL, U72933; AAB40699.1; JOINED.  
DR EMBL, U72934; AAB40699.1; JOINED.  
DR EMBL, U72935; AAB40699.1; JOINED.  
DR EMBL, U72936; AAB40699.1; JOINED.  
DR EMBL, U72937; AAB40699.1; JOINED.  
DR EMBL, U72938; AAB40699.1; JOINED.  
DR EMBL, U72939; AAB40699.1; JOINED.  
DR EMBL, U72940; AAB40699.1; JOINED.  
DR EMBL, U72941; AAB40699.1; JOINED.  
DR EMBL, U72942; AAB40699.1; JOINED.  
DR EMBL, U72943; AAB40699.1; JOINED.  
DR EMBL, U72944; AAB40699.1; JOINED.  
DR EMBL, U72945; AAB40699.1; JOINED.  
DR EMBL, U72946; AAB40699.1; JOINED.  
DR EMBL, U72947; AAB40699.1; JOINED.  
DR EMBL, U72948; AAB40699.1; JOINED.  
DR EMBL, U72949; AAB40699.1; JOINED.  
DR EMBL, U72950; AAB40699.1; JOINED.  
DR EMBL, U72951; AAB40699.1; JOINED.  
DR EMBL, U72952; AAB40699.1; JOINED.  
DR EMBL, U72953; AAB40699.1; JOINED.  
DR EMBL, U72954; AAB40699.1; JOINED.  
DR EMBL, U72955; AAB40699.1; JOINED.  
DR EMBL, U72956; AAB40699.1; JOINED.  
DR EMBL, U72957; AAB40699.1; JOINED.  
DR EMBL, U72958; AAB40699.1; JOINED.  
DR EMBL, U72959; AAB40699.1; JOINED.  
DR EMBL, U72960; AAB40699.1; JOINED.  
DR EMBL, U72961; AAB40699.1; JOINED.  
DR EMBL, U72962; AAB40699.1; JOINED.  
DR EMBL, U72963; AAB40699.1; JOINED.  
DR EMBL, U72964; AAB40699.1; JOINED.  
DR EMBL, U72965; AAB40699.1; JOINED.  
DR EMBL, U72966; AAB40699.1; JOINED.  
DR EMBL, U72967; AAB40699.1; JOINED.  
DR EMBL, U72968; AAB40699.1; JOINED.  
DR EMBL, U72969; AAB40699.1; JOINED.  
DR EMBL, U72970; AAB40699.1; JOINED.  
DR EMBL, U72971; AAB40699.1; JOINED.  
DR EMBL, U72972; AAB40699.1; JOINED.  
DR EMBL, U72973; AAB40699.1; JOINED.  
DR EMBL, U72974; AAB40699.1; JOINED.  
DR EMBL, U72975; AAB40699.1; JOINED.  
DR EMBL, U72976; AAB40699.1; JOINED.  
DR EMBL, U72977; AAB40699.1; JOINED.  
DR EMBL, U72978; AAB40699.1; JOINED.  
DR EMBL, U72979; AAB40699.1; JOINED.  
DR EMBL, U72980; AAB40699.1; JOINED.  
DR EMBL, U72981; AAB40699.1; JOINED.  
DR EMBL, U72982; AAB40699.1; JOINED.  
DR EMBL, U72983; AAB40699.1; JOINED.  
DR EMBL, U72984; AAB40699.1; JOINED.  
DR EMBL, U72985; AAB40699.1; JOINED.  
DR EMBL, U72986; AAB40699.1; JOINED.  
DR EMBL, U72987; AAB40699.1; JOINED.  
DR EMBL, U72988; AAB40699.1; JOINED.  
DR EMBL, U72989; AAB40699.1; JOINED.  
DR EMBL, U72990; AAB40699.1; JOINED.  
DR EMBL, U72991; AAB40699.1; JOINED.  
DR EMBL, U72992; AAB40699.1; JOINED.



```

OY 1741 QADEPFDSTIIQTIPFGVAL---ALGSIAFLF-----LKKTKASVGN----- 1781
Db 1423 DEDEETKEPELVQVHRNMVYIKLPHQVDGVQFMWDCCESVKKTKKSPGSGCIIAHQMGIG 1482
OY 1782 -----LFOILOIPKSDYDIPTLKSS-NRYIPYVSDRKGKTYIYMGDSDEDEKYAFM 1832
Db 1483 KTLQVVSFLHVLCDKIDFSTALVVCPLNTALNMNNEFEK-----MOEGLKDEDEKLEYS 1537
OY 1833 SDTDTVTSESEY---EELIDINDIYVPGSPKYKTLLEVVLEPSSGNNTTASGRNTPSDFRN 1889
Db 1538 ELAFYKRPQERSYMLQRMQEDGVMIIIGYEMVRNL-----AQRNVKSRKLL 1584
OY 1890 DIQNGCI---PSSKITDNEMQNLKKEFISNMLQNPNDYPSGNSSTNTINTTISR 1945
Db 1585 EIFNKALVDPGDFVVCDE-----GHILKNEASAVSKAMNSIRSRRIILGTGP- 1633
OY 1946 HNVDNNTTMTMSRDMMEENTLPSIHGNTLYSGEESYVNVNV--NSMNDIPINRDNNVY 2003
Db 1634 -----LQNNLI-----EYHCNVNFIKENLIGSKEFRNRFIN 1665
OY 2004 SGIDLINDLSGKRPDIYDEVYLRKRENELF---GTENTKRTSTQNVAKTNSDPINH- 2058
Db 1666 ---PIONGGCADSTWVDY--RYMKKRAHILYEMLAGCVQRK---DYTALTKFLPPKHEY 1716
OY 2059 -----QLELFHKMLDR-----HRDMCEKWKNKEDI 2083
Db 1717 VLAVRMTSIOCKLYQYLDHLTGVGNNSEGRGKAGAKLFODFOMLSRIWTHPWCLQILDY 1776
OY 2084 LNKLEEMNKENINNSGKTYNSDNKPSHNHVLNTDYSIOI-DMDNPKTKNEITNMDTNOD 2142
Db 1777 ISKENKGYFDEDSMDEFIASDSD-----ETSMLSSDDYTKKKKKGKKGKKDSSSS 1827
OY 2143 KSTMDTIILDDLEKYN 2157
Db 1828 GSGSDNDVEYIKVMN 1842

```

Search completed: May 1, 2001, 13:08:48  
 Job time: 316 sec



THIS PAGE BLANK (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2001, 13:02:07 ; Search time 53.97 Seconds

(without alignments)  
2837.034 Million cell updates/sec

Title: US-09-508-967-1

Perfect score: 12100

Sequence: 1 MATSGSGGTQDEDAKHVLD.....VNKKKEFEFEYPISDIMNI 2228

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12065	99.7	2228	2	T14029
2	5074	41.9	2197	2	B71600
3	4901	40.5	3078	2	T28432
4	4817	39.8	2182	2	T28634
5	4711	38.9	2664	2	T28626
6	4084	33.8	2647	2	T28161
7	4073	33.7	2042	2	T18399
8	3989	33.0	3006	2	T28625
9	3937	32.5	2924	2	T18378
10	3297.5	27.3	1711	2	C71625
11	3192	26.4	2706	2	T28155
12	2859	24.5	1729	2	T18396
13	2863	23.7	2212	2	T28157
14	2668	22.0	3026	2	T28431
15	2518	20.8	2135	2	T14602
16	1319.5	10.9	431	2	F71600
17	782	6.5	440	2	E71625
18	418	5.5	1435	2	A57933
19	406	3.4	248	2	C71624
20	360	3.0	4550	2	T18440
21	339	2.8	1526	2	A45605
22	329.5	2.7	3394	2	T18501
23	321.5	2.7	3523	2	T18477
24	321	2.7	2441	2	D71623
25	318.5	2.6	2401	2	T28676
26	315	2.6	1070	2	T30848
27	310.5	2.6	1979	2	C71622
28	303	2.5	2269	2	T28677
29	301	2.5	2829	2	A42771

30	300.5	2.5	3724	2	T18427	hypothetical prote
31	287	2.4	1308	2	E71622	probable membrane
32	284	2.3	6642	2	T29757	protein UNC-89 - C
33	283.5	2.3	3844	2	T18402	asparagine/asparta
34	282.5	2.3	3507	2	T34513	hypothetical prote
35	279.5	2.3	1939	2	T18372	repeat organellar
36	277.5	2.3	1658	2	S55101	hypothetical prote
37	271	2.2	2485	1	H71621	serine/threonine-s
38	269.5	2.2	1631	1	SAZQK1	major mezozoite su
39	267.5	2.2	1726	2	A45948	major mezozoite su
40	265.5	2.2	1639	2	S05603	major mezozoite su
41	265	2.2	2657	2	T18497	hypothetical prote
42	263.5	2.2	1640	2	A24594	probable major sur
43	262.5	2.2	1045	2	T18373	erythrocyte bindin
44	262.5	2.2	1726	1	SAZQGM	major mezozoite su
45	262	2.2	3147	2	T18674	hypothetical prote

ALIGNMENTS

RESULT 1  
T14029  
variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)  
N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)  
C:Species: Plasmodium falciparum  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jun-2000  
C:Accession: T14029  
R:Chen, Q.; Barragan, A.; Fernandez, V.; Sundstrom, A.; Schlichterle, M.; Sahlen, A.  
J. Exp. Med. 187, 15-23, 1998  
A:Title: Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PfEM  
A:Reference number: Z17860; MUID:968080592  
A:Accession: T14029  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2228 <CHE>  
A:Cross-references: EMBL:AF003473; NID:g2961467; PID:g2961468; PIDN:AAC05730.1  
C:Genetics:  
A:Note: FCR3S1.2-vari1

Query Match	99.7%	Score 12065;	DB 2;	Length 2228;
Best Local Similarity	99.8%	Pred. No. 0;		
Matches 2223;	Conservative	0;	Mismatches 5;	Incls 0;
Gaps 0;				
QY	1	MATSGSGGTQDEDAKHVLD	EFQKYHDEHGEAKNYSELKSLA	SILGETAFYVKS 60
DB	1	MATSGSGGTQDEDAKHVLD	EFQKYHDEHGEAKNYSELKSLA	SILGETAFYVKS 60
QY	61	MOTESKYTELIEANSKRNFCKKDKGNDVRSVKEQAGYDNKKMKC	INGMTCAFFRLH 120	
DB	61	MOTESKYTELIEANSKRNFCKKDKGNDVRSVKEQAGYDNKKMKC	INGMTCAFFRLH 120	
QY	121	LNKNFPPNNNSDSSKAKHDLAEVCMMAKYEGESIKTHYPRYDSKYVGSDFPCKTMLAR	180	
DB	121	LNKNFPPNNNSDSSKAKHDLAEVCMMAKYEGESIKTHYPRYDSKYVGSDFPCKTMLAR	180	
QY	181	SPADIGDIRGDDLYLGNNKKKONGKETREKLEOKLKEIFPKIHDN	KDKAOKRYNND 240	
DB	181	SPADIGDIRGDDLYLGNNKKKONGKETREKLEOKLKEIFPKIHDN	KDKAOKRYNND 240	
QY	181	SPADIGDIRGDDLYLGNNKKKONGKETREKLEOKLKEIFPKIHDN	KDKAOKRYNND 240	
DB	181	SPADIGDIRGDDLYLGNNKKKONGKETREKLEOKLKEIFPKIHDN	KDKAOKRYNND 240	
QY	241	EDPNFYKLEDMWTANRETVMGAMTCSKELDSSYFRATCNDTGGGP	JOHNKCRCKDK 300	
DB	241	EDPNFYKLEDMWTANRETVMGAMTCSKELDSSYFRATCNDTGGGP	JOHNKCRCKDK 300	
QY	241	EDPNFYKLEDMWTANRETVMGAMTCSKELDSSYFRATCNDTGGGP	JOHNKCRCKDK 300	
DB	241	EDPNFYKLEDMWTANRETVMGAMTCSKELDSSYFRATCNDTGGGP	JOHNKCRCKDK 300	
QY	301	GANAGKPKAGDGDVTVTPYFDYPOYLPMFEEMADEPCRRKKKLE	TEKCCRCCKDSD 360	
DB	301	GANAGKPKAGDGDVTVTPYFDYPOYLPMFEEMADEPCRRKKKLE	TEKCCRCCKDSD 360	
QY	301	GANAGKPKAGDGDVTVTPYFDYPOYLPMFEEMADEPCRRKKKLE	TEKCCRCCKDSD 360	
DB	301	GANAGKPKAGDGDVTVTPYFDYPOYLPMFEEMADEPCRRKKKLE	TEKCCRCCKDSD 360	
QY	361	ERYYSRNGYDEQFTSRKGVKMGKGTDFACGSYENIDNORK	JFDOKKYYTKEIS 420	
DB	361	ERYYSRNGYDEQFTSRKGVKMGKGTDFACGSYENIDNORK	JFDOKKYYTKEIS 420	
QY	361	ERYYSRNGYDEQFTSRKGVKMGKGTDFACGSYENIDNORK	JFDOKKYYTKEIS 420	
DB	361	ERYYSRNGYDEQFTSRKGVKMGKGTDFACGSYENIDNORK	JFDOKKYYTKEIS 420	

Qy	421	DGGGRKRAVGSGT	KYBGEYK	SPYELKJNDG	YTVQA	FLGLLN	NKACKD	ITDGGK	1MFK	480																																					
Dd	421	DGGGRKRAVGSGT	KYBGEYK	SPYELKJNDG	YTVQA	FLGLLN	NKACKD	ITDGGK	1MFK	480																																					
Qy	481	EVNSGGGVYGGSG	SGTSGAGST	IDENKGT	FEYRS	EYQCP	PDG	GYO	HKG	GNOM	ERTKTYK	540																																			
Dd	481	EVNSGGGVYGGSG	SGTSGAGST	IDENKGT	FEYRS	EYQCP	PDG	GYO	HKG	GNOM	ERTKTYK	540																																			
Qy	541	MRWSKLRPI	NGKVVLLSKL	KVYKDM	MLIK	NKWE	FECLT	ONS	SDGS	VGSV	VYTTG	ASG	600																																		
Dd	541	MRWSKLRPI	NGKVVLLSKL	KVYKDM	MLIK	NKWE	FECLT	ONS	SDGS	VGSV	VYTTG	ASG	600																																		
Qy	601	SEKKEI	DEMCKY	HNHNVQ	VVNO	GEVEED	DDCLK	AGGLC	ITL	PPK	NKNEV	SEAK	SGNN	660																																	
Dd	601	SEKKEI	DEMCKY	HNHNVQ	VVNO	GEVEED	DDCLK	AGGLC	ITL	PPK	NKNEV	SEAK	SGNN	660																																	
Qy	661	HADIOKTF	HDFEYFWVA	NHMLKDS	IHNRT	KRYK	CSIDG	TYMK	CRNG	CNNK	CC	FEK	WVQ	720																																	
Dd	661	HADIOKTF	HDFEYFWVA	NHMLKDS	IHNRT	KRYK	CSIDG	TYMK	CRNG	CNNK	CC	FEK	WVQ	720																																	
Qy	721	KETEMK	PIKDHFKT	OES	IPEGY	FTTL	LELIL	KLOFL	KEDTE	EN	TENS	IDA	EAE	ELKHLQ	780																																
Dd	721	KETEMK	PIKDHFKT	OES	IPEGY	FTTL	LELIL	KLOFL	KEDTE	EN	TENS	IDA	EAE	ELKHLQ	780																																
Qy	781	KILLEN	NNNLAAVNA	NACT	BECKT	LMDL	NH	LELND	AKC	DC	CLP	PRED	S	RGR	SAD	PSDI	840																														
Dd	781	KILLEN	NNNLAAVNA	NACT	BECKT	LMDL	NH	LELND	AKC	DC	CLP	PRED	S	RGR	SAD	PSDI	840																														
Qy	841	FIPPEE	KEDDEN	EDDED	ED	VED	RODE	ETAK	ETTE	SG	ATD	TTT	SLD	CP	IY	KVLT	FD	NSL	900																												
Dd	841	FIPPEE	KEDDEN	EDDED	ED	VED	RODE	ETAK	ETTE	SG	ATD	TTT	SLD	CP	IY	KVLT	FD	NSL	900																												
Qy	901	QDACS	LKYG	GNNS	FLGN	RCYT	PSGE	PTT	SSDK	NGA	ICVP	RRRR	LY	IKK	IY	DMAT	KTESP	960																													
Dd	901	QDACS	LKYG	GNNS	FLGN	RCYT	PSGE	PTT	SSDK	NGA	ICVP	RRRR	LY	IKK	IY	DMAT	KTESP	960																													
Qy	961	QASGE	ASST	SGST	TP	PD	SK	EALL	KAF	VE	SA	LET	FT	LH	NR	K	EEK	KAA	VO	B	G	A	G	H	GL	PIR	1020																				
Dd	961	QASGE	ASST	SGST	TP	PD	SK	EALL	KAF	VE	SA	LET	FT	LH	NR	K	EEK	KAA	VO	B	G	A	G	H	GL	PIR	1020																				
Qy	1021	VEBGS	PEYD	PE	DK	LE	KE	I	PD	GL	F	L	OM	F	Y	T	L	G	D	P	I	L	F	S	G	N	D	T	T	S	K	D	P	S	S	N	L	1080									
Dd	1021	VEBGS	PEYD	PE	DK	LE	KE	I	PD	GL	F	L	OM	F	Y	T	L	G	D	P	I	L	F	S	G	N	D	T	T	S	K	D	P	S	S	N	L	1080									
Qy	1081	KNI	VLLAS	GST	EO	ER	EM	K	N	K	Y	K	E	I	K	N	F	R	C	S	T	E	R	S	A	P	N	L	V	S	H	P	O	T	W	E	N	G	K	Y	I	W	H	G	W	C	1140
Dd	1081	KNI	VLLAS	GST	EO	ER	EM	K	N	K	Y	K	E	I	K	N	F	R	C	S	T	E	R	S	A	P	N	L	V	S	H	P	O	T	W	E	N	G	K	Y	I	W	H	G	W	C	1140
Qy	1141	ALTSK	DIT	AK	GV	EEK	POK	I	N	P	E	N	L	N	D	E	A	N	K	P	P	O	Y	O	T	Y	T	V	K	I	D	E	N	S	G	S	P	R	T	O	Q	A	1200				
Dd	1141	ALTSK	DIT	AK	GV	EEK	POK	I	N	P	E	N	L	N	D	E	A	N	K	P	P	O	Y	O	T	Y	T	V	K	I	D	E	N	S	G	S	P	R	T	O	Q	A	1200				
Qy	1201	SSDNT	PTTL	TF	HF	YV	R	PPY	F	M	F	E	M	E	S	C																															

Db	Accession	Protein Name	Length
Db	1501	K1SHCIKKGEGSKCINGCEKSKSCKLEKNIKKIAEMWIKKRFNDQYENKQDPDYNNKSI	1560
Qy	1561	LEELIPKIAVYVDDQDNVATKLCVFENSKGCTLLISNTQNNKENDADIDCMKLGAVAKNCPG	1620
Db	1561	LEELIPKIAVYVDDQDNVATKLCVFENSKGCTLLISNTQNNKENDADIDCMKLGAVAKNCPG	1620
Qy	1621	KPSGSKQSDCKEPPLPDBEDQNPENLTLEPPKRCPPPTTOPPEEKGGETCGNKEKKKDEK	1660
Db	1621	KPSGSKQSDCKEPPLPDBEDQNPENLTLEPPKRCPPPTTOPPEEKGGETCGNKEKKKDEK	1660
Qy	1681	KEESEPPAKKEEGSPAEPAPAESEETETNPPEPGGPAAPSTPAPPPDPPLPAP	1740
Db	1681	KEESEPPAKKEEGSPAEPAPAESEETETNPPEPGGPAAPSTPAPPPDPPLPAP	1740
Qy	1741	QADEPFDSTLIQTTIPFGVALALGSIAFLFKKTKKASVGNLFQILQIPKSDYDIPPLKS	1800
Db	1741	QADEPFDSTLIQTTIPFGVALALGSIAFLFKKTKKASVGNLFQILQIPKSDYDIPPLKS	1800
Qy	1801	SNRYIPYVSDRKKGKTYIYMEGSDDEKFAFMSDTDTVTSESEYEELINDIYVPSPK	1860
Db	1801	SNRYIPYVSDRKKGKTYIYMEGSDDEKFAFMSDTDTVTSESEYEELINDIYVPSPK	1860
Qy	1861	YKTLIEVLEPESGNTTASGKTPSDTRNDIQNDGIPPSKITDNEMNLKKEEISNMLQN	1920
Db	1861	YKTLIEVLEPESGNTTASGKTPSDTRNDIQNDGIPPSKITDNEMNLKKEEISNMLQN	1920
Qy	1921	QPNVDPNDYTGSSNSTNTNTTTSRHVNDNTNTTMSHDNMEENLLPSIHGDLNYSCEE	1980
Db	1921	QPNVDPNDYTGSSNSTNTNTTTSRHVNDNTNTTMSHDNMEENLLPSIHGDLNYSCEE	1980
Qy	1981	YSYNNMNMNSMDPIPNNDNNVSGIDILINSLSGKPIDIYDEFLKKEEELFETENTK	2040
Db	1981	YSYNNMNMNSMDPIPNNDNNVSGIDILINSLSGKPIDIYDEFLKKEEELFETENTK	2040
Qy	2041	RSTQNVAKKTNSDPIHNOLELFHKMLDRHRDMCEKMKNKEDILNKLEENNKENINNSG	2100
Db	2041	RSTQNVAKKTNSDPIHNOLELFHKMLDRHRDMCEKMKNKEDILNKLEENNKENINNSG	2100
Qy	2101	KTYNSDNKPSHHVYLVNTQVSIQIDMDNPKTYNETTINMDTNDOKSTMDIILDDLEKYNDPY	2160
Db	2101	KTYNSDNKPSHHVYLVNTQVSIQIDMDNPKTYNETTINMDTNDOKSTMDIILDDLEKYNDPY	2160
Qy	2161	YDFPEDDIYHNDVYEKSSMDIYVDHNNTYSNMNDVPTKMHITEMINVNKKKEFEDEY	2220
Db	2161	YDFPEDDIYHNDVYEKSSMDIYVDHNNTYSNMNDVPTKMHITEMINVNKKKEFEDEY	2220
Qy	2221	PISDIWNI 2228	
Db	2221	PISDIWNI 2228	

RESULT 2

variant-specific surface protein 1 homolog PFBI055c - malaria parasite (Plasmodium falciparum)

Accession: B71600

Alternate names: erythrocyte membrane binding protein 1 (EMP1)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000

C:Accession: B71600

R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Arai, L.; Koonin, E.; Parta, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.

Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743

A:Accession: B71600

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2197 <GAR>

A:Cross-references: GB:AE001434; GB:AE001362; NID:g3845341; PIDN:AAC71996.1; PID:g3845341

A:Experimental source: clone 3D7

C:Genetics:

C:Gene: PFBI055c



Db 1932 KDVBNDYKSGIPENTQ-----DNTLYFDKPEEKPEITSHIDNLLNGEYS 1978  
QY 1983 YNVNM-VNSMNDIPINRONNYYSGDILNDLSGGKPRIDYDEVILKRENELEFCGENTKR 2041  
Db 1979 YNVNMSTNSMD-PRYSNNYSGDILNDLSGKRIHIDYDEVILKRENELEFCGENTKR 2037  
QY 2042 TSTONVAKTNSDPIHNOLEFHMKLDRHDMCKEKKNEKEDILNKLKEMKKNINNSGK 2101  
Db 2038 TSHSVAKNTSDPIILNOINLFHTWLDHRDMCKEKKNEKEDILNKLKEMKKNINNSGK 2095  
QY 2102 TYNSDNKPSNNHVLNTDYSIOIDMDNPKETKNEITNMOTNODKSTMDTILDLLEK-YNDPY 2160  
Db 2096 THPSDS-----NKLNTDYSIOIDMDNPKETKNEITNMOTNODKSTMDTILDLLEK-YNDPY 2142  
QY 2161 YYYEDYEDIIYHDYDVEKSSMDIYVDH--NNVTSNNMDVPTKMHIEKNITVNNKKEIFE 2218  
Db 2143 YYDMYDD-YYDVN-----DHDSTVDTNADVPKVOIEKD-VNTR-LVKE 2187  
QY 2219 EYPISDIMNI 2228  
Db 2188 KYPIADYMDI 2197

RESULT 3  
T28432  
variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)  
N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)  
C:Species: Plasmodium falciparum  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
C:Accession: T28432  
R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson, D.S.  
Cell 82, 89-100, 1995  
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence at  
A:Reference number: Z20487; MUID:95330813  
A:Accession: T28432  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3078 <SU>  
A:Cross-references: EMBL:LA0608; NID:9886374; PID:9886375; PIDN:AA75396.1  
A:Genetics: var-1  
A:Gene: var-1  
A:Introns: 2611/3

Query Match 40 58; Score 4901; DB 2; Length 3078;  
Best Local Similarity 36.68; Pred. No. 1.2e-214;  
Matches 1174; Conservative 277; Mismatches 631; Indels 1128; Gaps 83;

QY 4 SGGSGTQDEDAKHVLDFFGQKVHDEVHGEAKNVSELSGLSLASITLGETAFTVKSMT 63  
Db 12 AAGGDDIEDESAKHMFRIQKDYDKVEAKERKGGRLS-----EAKFEKNSDP 65  
QY 64 ES-----KYTELEANSKNPCK-----KDGKNDVDRFSYKQAGYDNKKMKC 107  
Db 66 QTPEDPCDLDHKYHTNTTNN-VINPCADRSVRSDEYGCSTHNRIKISQOQDNKG-- 121  
QY 108 SNGTCAPFRRLHCNKNFPMNSNDSSKAKHDLAEVCAAAKKEGSIKTHYKYSKY 167  
Db 122 ---ACAPYRRLHVCNDLNLQIEPIKITNT-HNLLVDCMAAKFEGOSITODYPRKYATY 176  
QY 168 PGSDPFCMLARFADIGDIIRGRDYLGNKKKONGKETEREKLBOKEIFKFIHDN 227  
Db 177 GDSFSQCTMLARFADIGDVRGRDYLGNPQIK-----QKQLENNLKITGKIYEK 231  
QY 228 LKDEAKQRYNGDEDPNFYKLRDMWTANRETVGAMTCSKEIDNSSYFRATCNDTQGP 287  
Db 232 LKNGAE--RYG--NDPEFKLRDMWTANRETVKAITCNAM--GNIFYPATC---RG- 281  
QY 288 SGNHNRCKDKGANKGKRAAGDVTIPTYFDYPOYLRLWEEMAEDFCRKKKKLE 347  
Db 282 ERTKGYCRNDQD-----VPTYFDYVPOYLRLWEEMAEDFCRKKKKIK 325

QY 348 NLEKQCRGKDKSDEYRYCSRNGYDCEQTSIRKKGVRMKGCTDCEFAAGSYENNIDNRK 407  
Db 326 DVKRNCRGKDKEDRDCRSRNGYDCEQTKRALIGRLRGKQICSLYACNPVVDIINNKE 385  
QY 408 QPDKQ-KYTYET-----SDGGKRRKAVG--TTKEGYEKSFYEKLAKDGYDAF 458  
Db 386 QPDKQKKYDEIKRYENGAGSGGSRKRDAGTTTNNVDGYEKKEFYDELNKESEYRYDKF 445  
QY 459 LGLINNEKACDITD--GKIFNFEVNSGGVVGSGGSGTSGASTDENKGTFRYSRYC 516  
Db 446 LEKLSNEETCTKVDDESGTIDFKNVN-----SDTSISGASGIVNESGTYRSKYC 496  
QY 517 QPCPDGVO--HKG--NOMERKTRVKRMWSLYKP---INGKRVLLSKLAVVKDM 569  
Db 497 QPCPYCGYKAVNNGSSNEMEEKNN-GKCKSGKILYEPKPRDEGTTITLKGSKHND-- 552  
QY 570 LKKMKKEFLTONSSDGSVGVTTGASGSGSEKELDYDKCYKHNHVKOVVNOGEYEE 629  
Db 553 IEKLNKFCDEKNDJTNSSGSGTGGSGSGSGRQELYEEKCYKGBDVVVGHDEDEE 612  
QY 630 DDELKGAGGLCTLPNPKKNEVESEAKSQNNHADIQTFHDFEYVVAHMLKDSILHRTK 689  
Db 613 DYENKNAAGGLCTILKNQKKNEBEGANTSEKPEDEIQTFNFFYVVAHMLKDSILHRTK 671  
QY 690 RLKSCISDQKTMKC-RNGCNKKCDCEKWKYKQKETEMKPIKDHFKTQEGIPGYFTTLE 748  
Db 672 KLQCLQGNKRIKCGNKNKCNNDCECFKRWITQKDEMOKIYQHRKTQNIKRGSDMTAE 731  
QY 749 LI-----LKQFLKED---TEENTENSLDAEAEELKHLKILKLENNLAV 793  
Db 732 LIPDHVDVLYQNLQEEFLAKDSSEASEKSENSLDAEAEELKHLREITSESDNOEAS 791  
QY 794 VNAG-TEOKTLMDLKLNHLELDATCKDCPRLEDDK----- 828  
Db 792 VGGVTEOKNIMDKLNTKEADQLCEIHDEBEKEKDCGNCEIEGENFRNPPSCGE 851  
QY 829 ----- 828  
Db 852 SGNKRYPVLANKVAYONHNHAKQTOLASRAGSALRGDISLAQFNKNGNSTLKGQICKIN 911  
QY 829 -----SRGNSADPS-----PDIFIP----- 843  
Db 912 ENYSNDSKSGSGPCTGDDGHGVMRTIGTEWSNIEGKQTSYKNVFLPRRHEMCTSN 971  
QY 844 ----- 843  
Db 972 LENIDVGSVTKNDKASHSLGDLVOLAAKTDAEILIKRYKQDNINIQLDPIQKQDEAMCR 1031  
QY 844 ----- 843  
Db 1032 AVRYSPADLDGIIRGRDWMDEKSTDMETRLITVFNKIKKHDGINDNPKYTGDESKP 1091  
QY 844 ----- 843  
Db 1092 AYKRLRADWMEANRHOVRAWKCATGIIICPMFVDYIIPRLRMWTEMAEWYKASOE 1151  
QY 844 ----- 843  
Db 1152 YDKLKKICADCMKSGDKCTQGDVDCGKCAACDKYEEIEKMEQWRKISDKYNLLYLQ 1211  
QY 844 ----- 843  
Db 1212 AKTSTNPGRTVLGDDDPDYOQMVDFLPIPHKASIAARVLVYKRAAGSPTEIAAAAPTTPY 1271  
QY 844 ----- 843  
Db 1272 STAGYIHOEIGYGOEOTOFCSEKKNHATSTYTKENKEYTEFKORPREYATACDINRS 1331  
QY 844 ---RPEEKD----- 850  
Db 1332 QTEEPKKEENVESACKIVEKILGKNGRTTVEGCPNKESTPMDCKNNINIDISHGCMR 1391  
QY 851 -----DEN-----EDDEDEEVDR-- 863

Db	1392	PRROKCLLYIAHESQTEINIKTDONLKDAPFKTAAAEFTLSMOYTKSKSDSEAKILLDKGL	1451
Qy	864	-----DEETAKE-----TTGS-----	875
Db	1452	IPSOFLASMMYTTGDAYRDICLNTDISKONDYAKAKDKIGKFFNSDGSKSPSGLSROEWM	1511
Qy	876	-----ATD?-----	879
Db	1512	KTNBPEIWMGMLCALTKYVYVDTDNKRRIRKNDYSYDKVNSQNSGNPSLEFFAKPOFLBMM	1571
Qy	880	-----	879
Db	1572	IEMGEFCAEROKKENI1KDACNEINSTOOCNDAKHRNOACRAYOEYVENKKKEFGSOT	1631
Qy	880	-----	879
Db	1632	NNFLKANNVPOPEPKYKGYYKKGVOPIQGENEYLKOCNNKCSMDGNVLSVSKKEPF	1691
Qy	880	-----TSLDVCPIYGVKLYTKDNESL0DA	903
Db	1692	GKYAHKYPEKDCDYOGKHVPSIIPPPPPVPOPEAEPTVYDVSIV-KTLFMDTNNFSDA	1750
Qy	904	CSLKYGGNNSRLGRCY---TPSGEPTT---SDDKGAICVPPRRRRLYIKKIYDWMKTT	957
Db	1751	CGLYV-KCTAPSSKCIPLSDTKSGAGATTGKSQSDSGSICIPRRRRLYVKGIOEMATVL	1809
Qy	958	ESPASGEASSTSGTTPDPSKEALLKAFVESAL1EFPFLHRRKKEEKNAVAGACHG	1017
Db	1810	--POGECAPSHSA-----DDLNNFIQSALETETFLMDRKEEKPKPG-DG500A	1858
Qy	1018	LPRVEE--GSPEYDPEDK-LKEGKIPDGLFROFYLYGZYRILFSGSNDTYSVSKDPS	1074
Db	1859	LSQTLSTYSODEDEPPPKXLLQNKIPDFLRLMFTLYGZROILYHGN--TSDSGNTNG	1916
Qy	1075	SSNDLKNIVLLASGSTOEOREKMNKYKEI-KNFKCSTERSAPNLVSHPOTMENNKG	1132
Db	1917	SNNN--NIVLEASGNKEDMOKIOEKIEOILPKNGGTPLVPKSS--AQTPKMNNEHAE	1970
Qy	1133	YIWMGNCALT-----SKDKTAKVEKKPKOIEPBNLME-----ANKK-	1175
Db	1971	SIMGMICALTYTEKKNEDTISARDE--NTEEDDVEYKKEFGSTADKIGTASTPTGYK	2027
Qy	1176	POYQYTNVKLDENSQSPRTTOTQASSDNPTLTHFVRPRPYFRWFEEMGESFCREKK	1235
Db	2028	TQYVEYKVKLEIDSG-----AKTIPASSDTP-LISDFVLPRPYFLRLBEMGNPFCKKKKH	2081
Qy	1236	RLKOIKVDCKVEKNDVG-----RCSGDGEACDSI1SHDY1SVPSFNCPCGHRGHSY	1287
Db	2082	KLAQIKHECKYEEENGSGSRGIGTRQYSGDGEACNEMLPKNQJTVDPLEKPCACAPCSSY	2141
Qy	1288	RKWIERRKKIFHHQSNAYGQOQKTDARNNGNPTFDKEFCCTLLEWMPDAKLELRLKNGCK	1347
Db	2142	RKWLIESGKFEFEQEKAYEPOK-DKCVNSNKNKDNCFCLTLTSSAKAKOFLTL--GCKC	2198
Qy	1348	TNKEYGDDIDIFEKSDSTFOHTEYCGPCPRFKTQCNQNGCQVSLGNGNDGKGSIDAKEI	1407
Db	2199	PNNVEGTFID---DOKTFHHTDCOPCLKFSYNCKKDECD--NSKGTDCRKNKNSIADADI	2254
Qy	1408	AKMBSFTTDVVMRVSDNTFTFEGDDLKDACAQHANIFKGIKRDYMKCGYGVGVDICEQTN	1467
Db	2255	ENGVDSTIV-LEMRSADSKSGFNGDGLGNACGAGIFEGIRKDEWCKRNVGCVYVCKPEN	2313
Qy	1468	INEBTQKEYIQTIRALFKRVNVEFLDYKXINDKISHC1KKGBGSGCINGCKSNKCKLEK	1522
Db	2314	VNGEAKKH1IQRALVKRVNVEFFEDYNIKKIRISHRKNEIISPC1---KN--CYEK	2367
Qy	1528	WIEKKIAEWENIKRRFYDOYENKNDQPYNWS1LEELPKIAVWNOODVWIKLAFEWNSK	1587
Db	2368	WVDQKRKEMKEIITERFDQYKKNDSDDDNVRS1LETLFIQITIDAKAKNVIYLSKFGNSC	2422
Qy	1588	GCTLISNTQ--NNKENADIDCMKLTGVAKNCPK--PSEGEKSDCKEPPLEDE--	1639

[illegible]

[illegible]

QY	1054	YRDLFFSSNDTT--SVKOTPPSSNDMLKATIVLLAGSTPDEREKN-----KYREIKNFR	11089
Db	1020	YKDLIGSSNDISPTGTQRTPSSSNDMLKATIVLLAGSTPDEREKKKQIOAKIKILNGA	10799
QY	1109	KCSTERSAPNLYSNR--OTMMENNGKTYIMHAWCALMSKQIKAKGYEK--POKIEPNEMIL	11669
Db	1080	TSGVPPTKNSVKNRPQOTMMENLAKDIMNAMVACALTYKENDARGISAKIEQKKDKKALM	11399
QY	1167	DEANKPKRPQOYQATNVKLEBNGSTSPRTTOTQASSDNT--PTTLHPVKRPPTYRMEE	12249
Db	1140	DEANKNPETIEKYQTYNVKLEDES-----AKSNDTIOPTLKNFVEIPTFRMWLHE	11909
QY	1225	WGESCRERRRRKQIOYDCKVENGDYGRCSGGEGACDISITHDYSTVSEFCRCCGKHC	12849
Db	1191	WGNSECFERRARQIOIHNECDWDEGE--KOYSGGEGCEBFJSKOYVLODLS--SSCARPC	12489
QY	1285	SSYRWIERKKTEPHKOSNAYGOOKTD-----ATNNNGTPEKCEKTLFETMPRA	13399
Db	1249	RLYKWIETKKTEYEKOOKAYEOKKSAYNEBOKKOCOTOSNNNA--NEBSRLGASPTPA	13069
QY	1336	KFLLEKLNGPCKTKKEY--GGDD--IDPEKDSKPFQTEYEGCGRPKETKQCOGNAGVSGL	13929
Db	1307	BELOKL--GSCKNNGVENEGBDKIDPKNDKFKFAHSGDCRPIITGVACQNGHC--VESA	13659
QY	1393	NG-NCDGKSIDAKEIANKMSSTTDVYVYSDNDTFFE--GDCLKDACOHANIFKIRK	14499
Db	1364	NGECKKNK--TTAEDINKKTDPMNEMIVYSDSTYFELGID---CKSGGIFKIRK	14179
QY	1450	DYWKGVYCGVDIC--QOTMINBETDCKEYIOIALFKWVEVLEFDYKINDKISHQIC	15079
Db	1418	DEWKANVCGVDITLKKRKKINOEGDKITYIKELLKMYLEFLEDYNRIRKKIKLCTK	14779
QY	1508	KGEGSKJNGCEKNSKLEKWIETKILAEINIKRRPNDOYENKNDOPDYVNSILBELIPK	15679
Db	1478	KEDGCKCIKG-----CIEMKVOEKTEMKINDTYLEQYKND-----GNTLNFLEQ	15259
QY	1568	IAYVNDODNVKILC---VEENSKGCTLLSNQNKENDALDMLKILGVAKKNCGRPS	16239
Db	1526	FQYRTPEKNAIKPCDGLDOFKTSGILNSTQNGANNLDLVCLTLQKKISECKEQHS	15859
QY	1624	GKESQDC-----KEP---PRLPDEDDQNEBNTLEPKFCRPTTQPREBEGE--TCG	16719
Db	1586	GOTQPCPNSSLSGKESTLVEDDYDEEONP--BNKYEORFCRPMKEPKENDEVEGTCG	16449
QY	1672	NKEEKDKDESEEPKESGPAEAPAR--FAESELETN--FPEPCTGPAAPSPA	17289
Db	1645	GDEEK--KVEDSYIEQKEEBAASAPESPPLTPKAPKEENVKRP-----	16909
QY	1729	PTPTDTPP-----LRPADP--FDSTILOTTIPFGVALALGSIATFLKAKTKRASYGN	17819
Db	1691	-----PPPKRRIRIKTRNVLDHPAVIPALMSSSTIMMSIGFAAFYFLYKAKTKSSYGN	17449
QY	1782	LFQIQLPKSPDYDPTLKSSNRITPIYVSDYKAKKITYIHEGSD--EDKAFMSOTDVT	18409
Db	1745	LFQIQLPKSPDYDPTLKSSNRITPIYASDBHKKITYIHEGSGSGEKAFAFMSDTDTIS	18049
QY	1841	SESEYEELDINDIYVPSPPYKTLIEVLEPSSGNNTTASGKNTPSDTRDIOND--GIP	18979
Db	1805	SESEYEELDINDIYVPSPPYKTLIEVLEPS-----KDTQNDINDIPLSDIP	18539
QY	1898	SS---KITDENMNAKKELISNMLQNPNDVNDYTSGNSSTNTNTTSHNVNDNTN	19539
Db	1854	NSDPRPTTDENMQLLKDFISNMLQNTQTEPN-----ILHDVNDNTH	18989
QY	1954	TTMSGDMMENULLPSITHGONLYSGEYSYNV---NMVN-----SANDLPI	19969
Db	1899	PTMSHHNDQKPFMTSHDRNLNLFSGEYNYDNMNSGNPNINSIDSTNSMDSLTSSNHSPY	19589
QY	1997	NRDNNVYSGIDLINDLSGKKPLDIYDEVYKRENELEFOTENTKTRSTON--VAKTTSSD	20549
Db	1959	NKNDKULSGIDLINDALSNGH--IDIYDMLKRENELEFOTQHHKPNKITSNRKYVOTSSD	20179
QY	2055	PIHHQLELFFHKWLDNRHDMCEKKNKEDJLNLKJEMKNENJNNSGKYNSDNKPSHHV	21149





```
D 1507 IVLEASGNKEDMOKIOEKIEIILPRNGSTPLVPKSS---AOTPDKMNHEAESTWKGIC 1563
Q 1141 ALT-----SKDIANGVEKKPOKIEPNENLMD-----ANK-----PKPOYQATNV 1183
D 1564 ALTYTEKRPDIASNGDE---NKIEKDDEVYKFEFGSTADKIGTASTPTGYTKTQYDYEVK 1620
Q 1184 KLDENSGTSPPTTQOASDNTPTTLHFVKRPYFERNFEEMGESFCEKRRKRLKOIVD 1243
D 1621 KLEDTSG-----AKTPSASDTP-LISDFVLRLPYFRYLEMGONFCERKKRRLKOIKEE 1674
Q 1244 CKVENVGVCSCGSGEADCTSTHDYSIVVPKPCG---CGNHCSSYRGIWIERKKIEPHKO 1301
D 1675 CM--DGSCKKTSGBDEQD---RRDTSNVPSADLEGRCGSCSRYKWKIRKREYKO 1729
Q 1302 SNAYGOQKT---DATRNNG-NTFDEKFEKTLJETWPAKFLERLKNCGCKTNKEYGDDI 1357
D 1730 ANAYSKQTKYEEGSKGGLDHNKEFVKIGTCTDAAPLRLKNGCKKDNEGNGDI 1789
Q 1358 DFEKDSKTFQHTCYCGPCPKFKTCQNGC---GVSGLNGCDGKSIDAKEIAKRRSSTT 1415
D 1790 NFGNTEETFRPAENCKPCSSFKINCRCNGSCGDDTKEKNGG-TITGNFNTMGCTE 1848
Q 1416 DVNWRVSNQNTNTEEGDLKACQHANFKIRKDYMKCYVCYDICEQININERTDGK 1475
D 1849 DVNWHVSDKNANEEGDLDEACENAGIFGTIRKDEMKRCYVCGIHLICQKCKGNAIDQ 1908
Q 1476 EYIOIRALFKRWENFLEDYKINDKISHCIKKGSGKSCINGCEKNSCKLEKWKIEKIAE 1535
D 1909 QIILVRLALKRWVEYFLEDYKIKKIKLPCLENGSGTICNGCKKCRVEMWIKLAKDE 1968
Q 1536 WENIKKRENDQYENKDDPDYKVS---ILEELIPKIAVNDODNVIKLCVENSKGCTL 1591
D 1969 WTKIK---NHYLEKNKEGDKVNTSLVTWVLETLVQIAMAANDKREQTSLDKTKTSLGNC 2025
Q 1592 ISNQN---KNDALIDCKLKYKAKNCPG-----KPSSEKSDCKEPPRLDEED-Q 1642
D 2026 PENRKNDGNNDALIDCLNLETKIHCKQIHNSEVNSQPHNCGGNP-PDEEDL 2084
Q 1643 NPEENTLEPRKFCPTTPORPEEGGETGCKEKKKEKSESEBPAREBPAPART 1702
D 2085 LEENPVPQPGFCTPQOEPBP--DCKGKLEKKDEKEDKEPDEBPAREBDAI-----VPS 2138
Q 1703 AESEETETNPEPCTGPAAD-PSTAP-PTPPTPRLPQADEPFDSTILQTTIPFGV 1759
D 2139 G-----PGSEPEADKGPVKAPEIKPOEPDLSHVAIVP---SLVSTLAMSIV 2184
Q 1760 ALAAGSTAFLEKKTASVGNLFQIIOIPKSDVIDPLKSNKXIPVSDRYKGYIY 1819
D 2185 GIGFAALFYFLKKTSSVGNLFQIIOIPKSDVIDPLKSLPNKXIPYTSGKRYKRYI 2244
Q 1820 MEGDSDEDKYAFMSDITDVTSESEYEELINDIYVPGSPRYKTLIEVLEP----- 1871
D 2245 LEGSGSDS-GYTDHYSITISESEYEEMDINDIYVPSPKYKTLIEVLEPSCNNTTAS 2303
Q 1872 -----SGNNTTASGKNTPSDTRNDIONDGIPISSKITDNEWNOAKKEFISMLQNPNDV 1925
D 2304 GNNTTASGNNTTASGKNTPSDTONDIONDGIPISSKITDNEWNOAKKEFISQYLQSEPTYE 2363
Q 1926 PNDTSGNSSTNTNITTTSRNVNNTNTTMSRDNMEENLLPSTHOGNLXSGEYSINV 1965
D 2364 PN-----MLGYNVNNHTPTSSHNVKEKPIKMSIHONLFSGGEYNYDM 2408
Q 1986 -----NMVN-----SMDPIRNDNNYSGIDLINDSLSGKPIIDIYDEVLYK 2028
D 2409 FNSGNNTINISDTNSMSDLSNNHSPYNDKNDXLSGIDLINDALSQNH-IDYIDEMAKR 2467
Q 2029 KENLFCTE-NTKRTSTQNVAKTTNSDPIHNOLELFHKMLDRADMCCKMKNKEDILNKL 2087
D 2468 KENLFCTKHTKHTNTYNNVAKPARDPITQIIMLFHKMLDRHDMCEKMKNNHERLPKL 2527
Q 2088 KEENKKNENINSKTYNSDKRPSHNVLNDVSTQIDMDNKTKTNEITNMOTNDOKSTMD 2147
D 2528 KEIM--ENEHSGDI--NSGIPSGNHVNLNDVSTQIDMDNKTKTNEITNMOTNDOKSTMD 2583
```

```
Q 2148 TILDDLEKYNDRPYDYFEDDIYHDVDEKSSMDIYVDHNNYTNMMDVPTKMHTEM 2207
D 2584 TILDDLEKYNDRPYDYFEDDIYHDVDEKSSMDIYVDHNNYTNMMDVPTKMHTEM 2643
Q 2208 IVNNKKEFEEEPISDIWNT 2228
D 2644 IVNNKKEFEEEPISDIWNT 2664

RESULT 6
T28161
hypothetical protein FC93-va111-1 - malaria parasite (Plasmodium falciparum) (fragme
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28161
R:Hernandez-Rivas, R.; Matted, D.; Sterkers, Y.; Peterson, D.S.; Wellem, T.E.; Scher
Mol. Cell. Biol. 17, 604-611, 1997
A:Title: Expressed var genes are found in Plasmodium falciparum subtelomeric regions.
A:Reference number: 220483; MUID:97154495
A:Accession: T28161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2647 <HER>
A:Cross-references: EMBL:U67959; MID:91794255; PID:91809295; PIDN:AAC47438.1
A:Experimental source: strain FCQ27/PNG
C:Genetics:
A:Introns: 2158/3
A:Note: FC93-va111-1

Query Match 33.8%; Score 4084; DB 2; Length 2647;
Best local similarity 35.6%; Pred. No. 1.3e-177;
Matches 1036; Conservative 288; Mismatches 632; Indels 954; Gaps 106;

Q 6 GSGGTQDEDAKVNLDIEGQKVNDEHGEAKNVSELGKSLASILETAI--TVKSMQT 63
D 5 GRGRT-NKSAKEVLEDEIGETIQKKAHSDADTFRSQIKGN-----FGEAKFYNGEIMQP 57
Q 64 ESKYTEL---LEAN---SKRNPCKKDGKNDVDFPSYKEQAGYNNKKMKCSNGM----- 111
D 58 NSKLCEDHTDITVNTDGHSPC---EGROTVPFP-----DDNNSQCTKNRDKOSVD 106
Q 112 ----TCAPFRRLHLCKNFKPMNNSN--DSSKAKIDLAEVMAKYGESIKTHPYKDS 165
D 107 NSVGACAPRYRLHLCSNLSIESIOTNNYDSSAKKNLILAEVYAKAFEGESIVKMYEDLGH 166
Q 166 KYPSGDEPMCTMLARSFADIGDITIRGDLVYGNKKKKONGKETEREKLEOKLKEIFKKIH 225
D 167 H---TTEGICFALARSFADIGDITIRGDLVYGNPQ-----ESARRKOLEDMLRKIFEKIY 218
Q 226 DNL-----KKKEKORRNGEDENPFYKLRBDMWTANRETVYGAMTCKSELDNYSYRAT 279
D 219 KELTSSRNGKNGAIEERYK-DGSGNYTKLRDWMNANRLDTWKAMIC-KAPNAPYFRNT 276
Q 280 CNDTGGGSPQTHNRCRCDDKDGANAGRPKAGDGVTTVPTFYFDVVPQYLRWFEEBAEDFC 339
D 277 CSN---GEKPTGEGCQCI-----DG---TYPTMLDVLVPPQYLRWFEEBAEDFC 317
Q 340 RKKKKKLENEKQCRGDKSDERYGCSBNGYDCQOTISRKGVKVMGCTDCFPAGCSYE 399
D 318 RKRMLKIQNALKNCRGMDDDKEKCYCSRNGYDCTKTRISIDKYSMNRECTKCLVCDPVV 377
Q 400 NVIDOKRQPKOKK-----YTKIEDGGGRKKRAAGTTKYGC-YEKSFYELKKNDGY 452
D 378 KWINDKKKEFKOKKCCENELYRNNESSQSPK-----NTNNMETDYGMLKND-Y 428
Q 453 GTVDAPFGLILNKEKACDITDG--GKINFEKVNSSGGVVGSGGTSAGSGTNDENKSTFY 511
D 429 QSMNDFLKILNSETPCNIIDAKSIDF-----TKDDEE-TFS 465
Q 512 RSEYQCPQDGCQVQHKGNQWER-----KTKYKMKMSKLYKPIHG---KMYLLKSLK 562
```

Db 466 HTEYCDPCWCGIATQADGTWRLYENDPQCPKPK-----YEPKGVPEPTEDVLYTGK 520  
Qy 563 VYKDMIIILKKNKEFCILQNSDSGVSVYTTGASGNSEKKELEIDEMKCYHNEVQAKN 622  
Db 521 ENKDIIV---KLRECKT---DGNIGF-----KNE---EMWXY---YQVG 553  
Qy 623 VQGEYEEDDELKGGAGCILPKNPKNKEVSEAKSONNHADIQKTFHDFYFVVAHMLKD 682  
Db 554 NDKCVLENGEELGG-----EKVY-----KQYDNFLMVAHMLKD 588  
Qy 683 SIHMRTKRLKSC1-SDGKTMCRCNGCNKCDCEKFMVKOKETEMPKIDKHFTQEGSIPB- 740  
Db 589 SIEMRSK-LSNCKSDKKT--CITTCNDNCQCYDKWKIGKKVHMTQIKKHPDKQTFQGM 645  
Qy 741 GYFTTLEILKLOFLKEDTEENTENSIDAEAEELKLOKLEKLENNLAVNAGTGO 800  
Db 646 GHYF-VLETVL-----BEDQFTDITRAYGDAREIVHJOEMLOKKEEVLH--EDASNK 696  
Qy 801 KTLMDKLNHELNDATKC-----KDCPL-----PEED--KSGRSADPS----- 837  
Db 697 KTIIDELDLHELKAKOCIVNHNKDNCPADLSDSEDEEDIPQRONKCAKPSGTHIRALV 756  
Qy 838 ----- 837  
Db 757 NKVASNMHKKKQOLVNGVSSKLGDAKAGEYRKSGTTIKLKDICSITDHSNAKRGHT 816  
Qy 838 -----PDIFP----- 843  
Db 817 DOPCKRKDSKVNVKNRRWMDTAGFISNTYKIDYMPRRHOFTSNLEYLQTTNKLNGND 876  
Qy 844 ----- 843  
Db 877 INGNRIINDSLGIVLFRANTEADFIKMYKQNDYKDNATITCRAMKSFADLGDIDOR 936  
Qy 844 ----- 843  
Db 937 QHICRIMIVERVKEHSEIRNFLILSKNILAKFEIKEDTPYTKLREDWMEANRKKIMEA 996  
Qy 844 ----- 843  
Db 997 MOCPTPNSSPCKSYHIGLDYIPORLRMTWMAWFECKEOKKOYGEIVSASNGKDERV 1056  
Qy 844 ----- 843  
Db 1057 KYVLRVHNVQACRKHVITIKMLIHGKQMDKMEIKYLLYLOQTANGPDTYISGL 1116  
Qy 844 ----- 843  
Db 1117 VDENEKPVVNFLELYKENGCKIGNPRDTPRAKRSKRETA PASVAKNDVYSTAAGVHOE 1176  
Qy 844 -----RPEEKEDENEDDEDEV 861  
Db 1177 MGRHNECKTQTECEKRTDEQYENENTYFKNPPOYKADICNTRPPREKESRRKSEDSDE- 1235  
Qy 862 RODETAKET-DEGSAT-----DRTTSLDYCPITGKYLLXNESLODCSJK 907  
Db 1236 ---EEKVETKEVEKATEDAVDTGPPPAKKEATITLDVCPYIAGVLTJKN--LENCSPK 1290  
Qy 908 YCGNNSRLMGRCV-----TSGEPTTSSDKNGAICVPP 940  
Db 1291 Y-GRAPYTSWKIC1PPEKTMAATSGESSGNGALQRAKRAKYVSGSPVTSN--SGSICIPR 1347  
Qy 941 RRRRLYIKIIVMAT-----KTESPQASGSEASSTGTPPDSKEALLKAPVESAAI 993  
Db 1348 RRRRLYIQKLIHDMASGNTVVSQAOQTPQGG---TSSPSSKETPSPDK---LRTAFIQSAAI 1401  
Qy 994 ETEFLMHRKKEKKAQVAOE-----GAGHGILP--RYEBSPEYDEDEDKLKEK 1038  
Db 1402 ETEFLMDRKYKKAIAKAKKKKQOMVDSPLSTADPRHNNVSLVIAIPYIN-----KTCV 1456  
Qy 1039 IPDGLRQMFYTLGDYRDILFSGSNDTTSVSKDTPSSSDNMLKNITLLASGSTEOERERK 1098  
Db 1457 IPPEFLRQMFYTLGDYADLFF-GKND---IYIDTKNGKD-----IAERE--- 1497

Qy 1099 NKYKEIKNPKRCKSTERSAPNLVSHP-----OTWMEENCKYIMHGWCALTSKDKIANGV 1152  
Db 1498 ---KKIKD---AIEKVLKNAQSDQPSDEKROTWEOQNEHIMNGATCALTYREDEKGT 1550  
Qy 1153 EKKPOKIEBNMLMEANKPKRPQOYQYNNVKLDENSGSPPTTQOASDNT---PTTL 1209  
Db 1551 PLK-QNEGKLSALMDEKNNKPRADQKYQYKVLDENSGSPRIIVPAP:PTTTPPPPPPP 1609  
Qy 1210 THFVKRPTYFWMFEWGESCEKERRKRLQIYVCKVENDGVKCSGD:EACDSISTHDY 1269  
Db 1610 TSGF-SRPYFRYLEMAEFCEKERRKRLKIEKVECDDEGKKOKCSGD:EDEEIERKODY 1668  
Qy 1270 STVPSPGCGCKHCSYKMKLIERKKIEFHNSKSNYGOQKTATPANG:THPKECKTLE 1329  
Db 1669 STVRDQYCEGCKYCFRKYRGKKKDEYDKEKAEINNOKTARRNN--DNASTYTD 1725  
Qy 1330 TWPDAKFLERLKNGPCKTKNEYGGDDID-----PEKDSKTYQHTZYCPCPKFKTN 1381  
Db 1726 TCTTAGDFLQTLKNGPK-----NDNVDSGENKKIPENDDTFKYTJYCGTSLNGFK 1779  
Qy 1382 CQNGMGVSGLNGCDGK---SIDAKETAKRSSTTVYKRVSDNTYFEGDCLKDAC 1438  
Db 1780 CNGDCQVRV-TWVTCNGSNRTTITTAADI-KNGNSABEINMLVSD-DINSGNGFNDLEAC 1836  
Qy 1439 QHANTFKGIRKDVWKCQYVCGVDCIQEQTINERTDQKEIOTRALFKRVENLEDYNTI 1498  
Db 1837 KNANIFKGIKEKMKCVYFCKSDVGLKKNND-IDQNOIILLRALFKHLEFELDYNTI 1895  
Qy 1499 NDKISHCIKKGESKINCCEKNSKLEKMEIKETIAEMENIKKRFNDYKENDQDPDYNK 1558  
Db 1896 RKKLNPCINNGEKAJCTNG-----CVEQMINHKREWMNLKS-FNEC-NDODTERNR 1947  
Qy 1559 -SILEELPKIAVNDQNVYIKLVFENSKGCTLSNTQNNK---ENF AIDMLKLGK 1614  
Db 1948 LRFVLDLIRQIATIDKGNHGLVYKVSVCNCGNSONGKEGBENTLYLCLQKLEKK 2007  
Qy 1615 AKNCPGKP---SGEKSQDCKEPPRLDEEDQ--NPRENTLEPKFCP----- 1656  
Db 2008 AERKADNPETSGIPQOPCEVSPNHIJEDQPLBEEENYEHKICQDI LKHNHNOBER 2067  
Qy 1657 ---PTTOPPEKGETCCKNKEEKDEKKESEEPAKESGPAABEP:PTAASEETENF 1712  
Db 2068 LYKNPLVQPTLKRKKK--KKRKKIKKKKNQDHPHRLPCAFINTNP-----KTK-- 2118  
Qy 1713 PEPGTPAAPSTPAAPTPDTPRPLRQADEPFDTIIOTTIPRGVILALGSIAPFLTK 1772  
Db 2119 --PSSG-----KNPMEHPAIVP-----ALVTSILAMSV:IGFAAFYFYFLK 2158  
Qy 1773 KTKASVGNLFOILQPKSDYDIPPLTAKSNRYIPYVSDYKKTQYIYIEGSDDEKYLAM 1832  
Db 2159 KTKKSTIDLLSL--IPKSDYDIPKLSPNRYIPYISGKRGKRYIY:EGDSGTOS-GYT 2215  
Qy 1833 SDTTPVT--SSESEYELELDINDIYVPGSPKYKTLIEVLEP---SGNTYTAGSKN-TPSD 1886  
Db 2216 DRYSDITSSSEBYEEMDINDIYVPGSPKYKTLIEVLEP:SGKLSGNTP:PTSGNNTIAD 2275  
Qy 1887 TRNDIONDGIPSSKITYDNEMNQLKEFISNMLQONPDVNDYTSGN:STNTNITTTSH 1946  
Db 2276 TQNDIPTSDTP-PPTJDEDMNTLKHDFISNMLQONQKQDVNDYTSGN:STNTNITTTSHD 2334  
Qy 1947 NVDNNTNTYASDNMEENLLPSIHQNLYSGEBSYNNVNNVNSMNDIP:INDNNTVSGI 2006  
Db 2335 NVDNNTNHPYMSRHNNDQKPFITISHDRNLTYGEEVYNNVNNVNTJMDIP:INSHNNVSGI 2394  
Qy 2007 DLINDLSGKRPIDYDEVYLLKRENELFGEYNTKRTSTONVAKTYS:DPJHNOLELFHK 2066  
Db 2395 DLINDTSCNEHIDIYDELKRENELFGEYNTKRTSTONVAKTYS:DPJHNOLELFHK 2454  
Qy 2067 LDRHRDMCKMKNKEDILNKLKEENKE---NINN--SGKTYN--SDNK----- 2108  
Db 2455 LDRHRDMCKQKNDNERLAKLKELMENETQCGDINSIGSKLSPRSDNNIHSIDHPSD 2514



1429 PNOVGENQNF--GDSLPHVDDDDDEELETTEENTYVQRPICQLKRPQPQPG----- 1479  
QY 1675 EKKDEKKEESEEPANEESSPAEAPPAETAESEETETNPEPPOGTGPAAPSPAPPTPDY 1734  
Db 1480 -----DEDCKQASAPAP---SEGTEEN---QPPVIRKEEAPAPRPRPRPR 1518  
QY 1735 PPPLRQPADP--PDSITLQTTIPFGVALALGSIAPFLAKKTKKASVGLFQILQPKSD 1792  
Db 1519 KPPQEQPPLPPALKNAMLSSTIMSIGAFATFYFLKKSXSSVGLFQILQPKSD 1578  
QY 1793 YDPTLSSNRKIPYVSDRKGTYYLYMGDDEDEKFAFMSPTDPT--SSESEYEELDN 1851  
Db 1579 PPIPLKLSNRKIPYTSKGRYAKRATYYLEGDSDGDS--GTYDHSYDSTTSSESEYEEMDIN 1637  
QY 1852 DIYVPSKPKYKTLLEVLEP-----SGNNTTASGKNTPSDTRNDIQNDGIPSSKITDN 1904  
Db 1638 DIYVPGTPKTKTLLIEVLEPSCNNNTTASGNNTTASGKNTPSDTONDIPSGDTPNNKLTDN 1697  
QY 1905 EKNOLKKEEISMALON--QPNVYPNDYTSNGSSNTNTITTSKHNVDNNTNTMSRDNME 1963  
Db 1698 EKNNTLKDEFISMALQSEOPKDPNDYSSGDIPEPNTQ-----PSTLYEPDNQGE 1744  
QY 1964 NLLPSIHGNTLYSGEYENYNNM--VNSMNDIPINRDNVYSGIDLINDLSGKRPIDY 2022  
Db 1745 KPFITSIHRODLTGEYEVYNNMSTNSMDDIPISGKNVYSGIDILM--TVCGNHAVDLY 1803  
QY 2023 DEYLARKKENELFGTE--NFKRTSTQNVAKTTNSDPIHNOLEFLHKWLDLRHRCCEKKWKE 2081  
Db 1804 DELLRKKEENELFGTKHHTKTYVYVAKPARDDPLHNOLEFHTWLDLRHRCCEKKWKN 1863  
QY 2082 DLNLTAKKEEM-----NKENINNSGK-----TVNSDKRP----- 2109  
Db 1864 EKLAKKEEMENDTSGSKHSDIPSGKLSDPSPDNHSHSDIHPSDIPSGKLSDIPSDKY 1923  
QY 2110 ---SHNHLTVDSIQIDMNPKTKEI-----TNMDTNDKSTMDTILDDPK 2155  
Db 1924 CCSSNKTLTWDYSIQIDMNPKTTFNPTTYVDSNPNOVDYDTYDSNPNDSMPTILEDLKR 1983  
QY 2156 YNDPYVDFEYEDDIYHDVDEKSSMDIYVDH--NNVTSNNMDVPTKMHIEKNYNNRK 2213  
Db 1984 YNEP--YVVDLDD--LYVDVN-----DHDASTVDSNNMDVPSKVOIEND--VNTK- 2027  
QY 2214 EIFEPEYPISDIWN 2228  
Db 2028 -LVKEKYPIADLMDI 2041  
RESULT 8  
T28625  
variant-specific surface protein 3 - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
C:Accession: T28625  
R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herfelde, J.A.; Peterson, D.S.  
Cell 82, 89-100, 1995  
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence an  
A:Reference number: 220487; MUID:95330813  
A:Accession: T28625  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3006 <SU>  
A:Cross-References: EMBL:L40609; NID:g886376; PID:g886377; PIDN:AAA75397.1  
C:Genetics:  
A:Gene: var-3  
A:Introns: 2597/3

6 GGT-NTAKEVLDDEICKEVQEKATEBDALTYRNDLQNSQAQFHG--VP IDVAINPCDLNTE 63  
QY 65 SKYTELIEANSRNPCKDKGKNDVDRFS--VKQAGYDNKKMKCSNG-4-TCAPFRLHL 121  
Db 64 I--HTVNVKRRKKENPC-----RGREKRFSFDVLSGQCAKNTIDQSYTNS/GACAPFRLHV 118  
QY 122 CKNRFPNANSNDSKRAKHLLAEVCAAYEBSISKTHPKYDSKYPG--SDPFCMTILAS 181  
Db 119 CDRNL-ELIKTEBNTSTHDLVDVLTAKHGESLVKHHKEYIKK--NBNFICVYLAHS 175  
QY 182 FADIGDIIGRDLVYGNKKKKONGKEREKLEOKLEIFKTLIHDL-----KOK 231  
Db 176 FADIGDIYVKGDLVLYCYDDEEKE-----KKOLEKNUKKFFQIHDDVMTSGRTNKKSA 231  
QY 232 EAOKRYNGDEDPNFKYLRBDWMTANREYVWAGTCSKELDNSSYPATCN-----DTGOG 286  
Db 232 EAOKRYN--DATGNYYKALREDMNNANRDQYKAITCD--AADNDEYFENS SDGLYVFSNG- 288  
QY 287 PGTNKKCCGDKDKANACKPKAGGDVTVTPYTPYVYVYLYRMEPEKAEDECRKKKKL 346  
Db 289 -----C-----GRNECK-----VPTNLDYVPHLRWEDMAEDCCRKNITL 325  
QY 347 ENLEKOCRGKDKSDEYRYCSRNGYDCEOTISRKGVKMGKCTDPCFGSYENMTDNOR 406  
Db 326 KSAKEKCGEGNT---KYCSLNGYDCTKYFEKKDKSSDGNCTAGSNJCIAIDAMLNROR 382  
QY 407 KQFDKOK--KYTEISDGGGRKKRAVGGTTRYEG-----YER-SFEYELKNDGYG 453  
Db 383 NEFEKOKIYTFEIE-----KYSSSDKSNSNISKYHIEYENFGKKEYE 428  
QY 454 TYDAFLGILANNEKACDLTDGKINPKEVNSGGGVGGSGGTSAGS/YNDEKKGFFYS 513  
Db 429 TLQNFPLKLLNKMGYCOE-----KIEEEV-----IIPNKEDVYFHS 466  
QY 514 EYQCPPCDGYOHKG-----KIEEEV-----IIPNKEDVYFHS 529  
Db 467 EYQCPPCDGYOHKG-----KIEEEV-----IIPNKEDVYFHS 526  
QY 530 -----NOMERKTK 537  
Db 527 EYFCSSTYNEGRNVQKWKCYKNKSDNYNCMNISYSKSDTANVYLIVECFHSAKKNLL 586  
QY 538 VKKMRMS--KLIPYINGKMYLLKSLKYVD----- 566  
Db 587 IDTIKWEHQLKNCINNTNTYTCES--KCIKNCEYKEMIKRKEHEWEK/KNVEGNNNRMSY 645  
QY 567 -----MMIL-----KKMKKEF----- 577  
Db 646 IYYNNLSRVDFSEFLFOVMFALDODEKGWDOFTEDLKKRFPESKTNTPYTKSODALIEPLL 705  
QY 578 -----CLTQNSSDG--SVGSVYTTGASGENSEK-----KELYDEMKCYKHNHYOK 620  
Db 706 DHIKQDALCLCRNNSNESCDYKVKVYTNPCGNPSPASNNLVKRLAEEMQRYARQOLEK 765  
QY 621 ---VAVQGE----- 626  
Db 766 RGEELMLKGDATKGYRROGAPADGFEKNVCSINQNTNVONNRATYVJGPTCTGKDGSGNG 825  
QY 627 -----VEEDDEELGAGGICL----- 642  
Db 826 VRMKIGTTPMKPGROIOMSAEDIYMPRRONMCTSNLEVIQTYDGPLKGGDKLVNNSFLG 885  
QY 643 -----LPNPKKNKEV-----S 653  
Db 886 DYLLSAKMDAGKIIELYKKQNNKSNLTPDEDENSACRALRYSPADGLDIRGRDLMDKNS 945  
QY 654 EAKS-ONN-----HADIOK-----TFHDEFT--YVVA--HMLKDSIHMKTRK 990  
Db 946 DAKRLQTNLKEITFTKIKEELPEDIKKKYDKDTHKLLREDWMEANFHOV-----WRA-- 998  
QY 691 LKSCISDGKTMCRNG----- 706

Dh 999 MCALENDKDMKC-NGIDIEDYIPQRLRMWTEMAEMFCKEOSRLYNKLVADKSCCKGNK 1057  
Qy 707 -----CNKKCDCEKVVOKETEMKPIK----- 729  
Dh 1058 SCTOKDGCOTCKACACDYNKKIKIMEOWEKIKKKYAOQYKALDYNKGEESKKTAS 1117  
Qy 730 ----- 729  
Dh 1118 DAKDOVVHFLAELIRKSGGKCKGNVKTYSPTTTPNTLYSSAGYIHHELGRTVGCNT 1177  
Qy 730 ----- 729  
Dh 1178 QKEFCYKNGKAYAFKDPKGYEACCKNDRNPKQPARAKKEDEDACDVYKLLKDKGETD 1237  
Qy 730 ----- 729  
Dh 1238 DIDGCONYKAKGDKYPCGWDNCSQIHTHNGACMPPRQKCLCVSGLTJTDRIKAIETRT 1297  
Qy 730 -----DHFQTOEG-----IPRGY-----YFT----- 745  
Dh 1298 EFIKSAIETHFPAWDRYKEDNGEAELKNGINIEGFKRQWYTFGRDYRDIFFGNDISTH 1357  
Qy 746 -----TEELIKLOFLKEDTEENTENSL-----DAE 771  
Dh 1358 AYISGVSPKVTITIEKENDAYAAKQNSNELLDDMDHOGKDIEGMICALTHKTSDEE 1417  
Qy 772 EAELKHLQKILKL----- 785  
Dh 1418 KKEELKNYKTKLINESPKGSKNVDFPAKKPOLFWFIEMGDEFCAQREBEKAKYVSCS 1477  
Qy 786 -----ENENLAVNA-----GTEQKTLMDD--KLLNHE----- 811  
Dh 1478 DAKDYGCKNKNKSNASCYKVEDYITKKKVEYTKQKGFDAKIDTKEGYEGFSTKD 1537  
Qy 812 -----LNDATKCK-----DCPLBEE--DKSRGRS 833  
Dh 1538 ASEYLKKKCLDDTCNCMOKVKNNTYEWNTPNKTYTNSMLEKRCQEPQEPPEGGAR 1597  
Qy 834 ADPSHDIFPRP-----EKKEDDN----- 853  
Dh 1598 SDSGRDTPPRAGSDASNVPSPRRPADYHEVAEVOGEDEDEDGLPELODEDV 1657  
Qy 854 -----EDDEDEVAD----- 863  
Dh 1658 EVAGAEEDDLDVGVARILGRRTNSPDEDEDEASEEDDDDDADQDTTEVYGQGEETAE 1717  
Qy 864 -----DETAKEETBESA-----TDTTSLDVCPIYKVLTDNLSLODACLK 907  
Dh 1718 DHOTTEETVQEKAEEDKDGGETPQKETQPKVEVNPCLDV-KTLFTTTEELKACPTK 1776  
Qy 908 Y-GGNNSRLGRCVTPSGEPTTSSDKNGAICVPPRRRLYIKKIVDMATKTESPQASGE 966  
Dh 1777 YVNGEKEPRNMKCI-----SSGSDASGSIICPPRRRLYLAKI-----EGVD 1818  
Qy 967 AS-STSGSTTPPDSKEALLKAFVESAIETFFLHRYKEEKAAVQEGAGHG--LPRAVE 1022  
Dh 1819 TTVSSDGETTPITHDALREAFIQTAAVETFFLMHRYKKIKEKEQOEELQNGTFLPAPQ 1878  
Qy 1023 EGSPVEUDE--DKLKECKIDGFLROMFYTLGVDYRDLFLFGSNDTYSV--SKDTPSSN 1077  
Dh 1879 KVSPEDNENHOKLKECKITBEERKQMFYTLGDRDLGVGVDAQALESGDNKSGD 1938  
Qy 1078 DNLKNIYVLASGSTEOEREKKNKYKEIKNFKRCSTERSAP-----NLVSHQPTWMENNGK 1132  
Dh 1939 KNINDI-----SEKISYIEKSEGOJTPPGPKPQGTITKXBEEMQKNGE 1981  
Qy 1133 YIMHGMVCAIT--SKDKAKGYEKKPQKIENTENIDEANKKPKPPQOYQYTNVKLDENSGT 1191  
Dh 1982 HINNAMICALTHNTDROVDQVKGQLEFENGKN-----TPKNSQOYQYKNV----- 2026  
Qy 1192 SPRTTQOASDNNP--TTLTHFYKRTYFRWPEMGESGCRREKAKKLOIKDCKAYEN 1248  
Dh 2027 ----TIVSSVSSNGPGIGNIKLEOFASRPTPLRWLEWGEFCCRQOYHKLRIKECHKDG 2082

Qy 1249 GDVGRSGDGEACDSISTHDYSTVPSFNCPCGCKHSSRYKRIEERKIEFHKQSNAYGOQ 1308  
Dh 2083 N--RNCDDDGPECKCMCKCKGOSFETLKCPCSAKSKCYKWKWISKKKBEFLKQKAYIKQ 2140  
Qy 1309 KTDATRNNGNTFDEKFCCTLETPDPAKFLERLKNPGCKTNKEYG-GDDIDFEKDSKTFQ 1367  
Dh 2141 KKDAGNNNDY--KEFSKTLRNYNDAALFNLKLNKPCSKNDSDVODEIKFDDERKTFG 2198  
Qy 1368 HTEYCGPCPKREKTC--QNGNCGVSGLNCGDKSDIDAKELAKRSSTTDVVMKRSNDT 1426  
Dh 2199 HETTCPCSKITVCKKENHNDNSKN--DCRNINSISADEIKSNSNTQDVTMVSDBNT 2257  
Qy 1427 NTFEGDDLKDACOAHANIFKIRKIDWKGYCGVDIC--EOTNIEPRTDKE-----YI 1478  
Dh 2258 NGNKFTYDLND-CIRKAGITKGIREDYWKCGEYCGVDICTLEKTNNEERYSAKENDKNOIT 2316  
Qy 1479 QIRALFKRWENFLEDYKINDKISHCIRKGSCKINGCEKNSCKLEKWKIEKIAEMEN 1538  
Dh 2317 LIRVLFRWLBSFLBDYKINDKISHCKNDKSPCINGCQNKCNCEYKIEKKKSEWGK 2376  
Qy 1539 IKKRPNDQENKQDPD-YNVSISLEELPKIAYVNDQNVIKLVFENSKGTLSINTQ 1597  
Dh 2377 VRERYINQYRDKNSNEAEVKSFLETLIPQIPVYTDKGDHSLQKLKLCNCKSEKEN 2436  
Qy 1598 NKENDAIKMLKLGKVKAKNCPGKPSGKQSDKEPPLPEEDQNP-----EENTL 1649  
Dh 2437 SNEKDVIVCLKLKLEDKAKNCKQDASGE-----PCPOTTSENPDEDILLEEENPV 2487  
Qy 1650 EPPKFCPTTOPPEEKGETCGNKEEKKEKKESEEPKAEESGPAAEPATSEETE 1709  
Dh 2488 EAPNICPKVEEP-----EPVVEEKCDALEAPSKESS-----TEENSGEG 2527  
Qy 1710 TNPPEPGTGAAPSTPAPPTPDPPLRP-----QADEPFD--STLTOTTPRGVLA 1762  
Dh 2528 SNSQNPRSKPEEPPEPTTSSETDTPPAPPTIOPSQADQPTNSISDLSTIPGLAIA 2587  
Qy 1763 LGSIAFLKRRKTKYASVGNLFOIOLIPKSDYDITPLKSSNRYIYVSDYRGKTYIYMEG 1822  
Dh 2588 LTSIYFLFKKTKTSSV--DLRLVNLIPKGEYGMPLTKSSNRYIYASDRYKGYIYMEG 2646  
Qy 1823 DSDEDKVFMSDQTDVNTSSESEYEELDINDIYVGPSPKYKTLIEVLEPSSGNNNTTASGKN 1882  
Dh 2647 DSDSGHY--YEDTJDVNTSSESEYEELDINDIYVGPSPKYKTLIEVLEPSSGNNNTTASGKN 2704  
Qy 1883 TSPOTRNDIONDGIPTSKITTNENMOLKEFISNMLOL--QNDVPNQYTSGNSSNTNIT 1941  
Dh 2705 TSPOTNDIONDGIPTSKITTNENMOLKEFISNTLOEIOSEOPNDVNDTSGNSSNTNIT 2764  
Qy 1942 TTSRHNVNNTNTTMSRDNMEENLLPISHDGNLYSGEESYANNVMSNMDIPINRDN 2001  
Dh 2765 TTSRHNV-----BEKPFINSIHDRNLTYGGEIYNNVNMVMTMDIINRDN 2811  
Qy 2002 VYSGIDLINDLSGSKPIDIYDEVYKREKENLFEGETNKTSTONVAKTNSDPIHNOLE 2061  
Dh 2812 VYSGIDLINDLALNG--DVDIYDEVYKREKENLFGFNHVKQTSISHSVAKPARPDHINOLE 2869  
Qy 2062 LFHKMLDRHDMCKEKKKEDILNKIKREKKNENINNSGKYNSDNKPSHNVINTQVSI 2121  
Dh 2870 LFHKMLDRHDMCKEKKKEDILNKIKREKKNENINNSGKYNSDNKPSHNVINTQVSI 2121  
Qy 2122 QIDMDNPKTNEITNMDTNDQKSTMDTILDLLEKYNBPYVYDFEEDIIYHDVDEKSSM 2181  
Dh 2924 QIHDNPKRINIEFIN-----MDTILDLDDLINNE--YDVQDD--IYDVAN----- 2965  
Qy 2182 DDIYVDH--NNVTSNNDVPTKMHIEBNIVNNKKEIFEERYPISDIWINI 2228  
Dh 2966 ----DHDTSTVDNSNADVPKSVQIEMD-VNTK--LVKEKYPIDVYDI 3006  
RESULT 9  
T18378  
variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)





```

Oy 1569 ----- 1568
Db 1711 TYENSVNINNKCKDNONEFKIGOKMNFYIGIRKOLCIPPREHMCDDLMLGRTT 1770
Oy 1569 -----AVNDQDNVILCVFENS----- 1566
Db 1771 ISDSALLKRIQEAQKSRBDIIRKLEONSCDEHRICDAMKYSFADIGDIIRGRDLNKK 1830
Oy 1587 -----KGC-----TLISNTONKKNEND----- 1602
Db 1831 NSKOKGLOKLEAFINITYNKLQNDKNKYEKDREKYLQLRSDMWDANKHIMAMTCNAP 1890
Oy 1603 ----- 1602
Db 1891 DDAFKLKNPNDTSGSSSKGIMTTHSNCGYDKPEPDYDIPQPFMMQEMSSEFCRLN 1950
Oy 1603 -----AIDCKLAKLYKAKKPC----- 1620
Db 1951 EEMQEPKTCGECKKNSITCEDDRNGTNCENCKNCKEYKKLHNMWKLGFDKYEIYNEI 2010
Oy 1621 -----KPSGEKOSD-CKE----- 1632
Db 2011 YNNKDSKINSNEYKFKLEKLDKCKELNSDKCIDENATHTCTKTKFSNSEKNHNNNTAFK 2070
Oy 1633 -PP-----PLPDEEDONPEE-----NTLEPPKFC----- 1655
Db 2071 NPPEYKACKCDAPDPLDNCPKDSATYERKACNTLPTKLCESKTPFNNDSDSMTSFYQT 2130
Oy 1656 -----PP----- 1657
Db 2131 SPRDNTGVLVPPRRROICLKNITTKLSIEKIDPFKALMTSAYNEGKLLCELYKKDRDV 2190
Oy 1658 -----TTOPPEKGG----- 1667
Db 2191 TLQAMKSYDYGDIVGTDLISTAPLDKLTKLNLVLLKGDTMEIKEDRCKMWTENRTR 2250
Oy 1668 -----ETCGNKEKKD-----BKKESEEPKAEESG-- 1693
Db 2251 VNHAMLCGYKKAQKIEERDCSLPDNDHQLRFWRFEWSEHFCAKROGLFNEVREKASA 2310
Oy 1694 -----PAEEFPAP-----TAESEB-----TENPEPPECTGAP-- 1723
Db 2311 QCILEYGTIDPPVCEACTOYRDYITRKIOEYRLNLYOYNENFNEKKAETKAEYENDK 2370
Oy 1724 -----PSTP-----APPTD 1733
Db 2371 CNOKCNLSKYIDIEKKKMNMYDSFDNDLKNKICIQIKRKRPPKVKYKPEEHTPESQD 2430
Oy 1734 TPEPLRPQ-----ADEPEDSTLQTTIPGVALALGSIATFLKKTAKSVGNLFQIL 1786
Db 2431 TPEPLRPQDLPPEAPEEPRDILEKTIIPGIALALGSIATFLKKTAKSVGNLFQIL 2490
Oy 1787 QIPESDQDIPTKLSNRYIPVSDRYKGYIYMEGSDDEKAFMSDTTQV--SSSEY 1845
Db 2491 HIRPSDQDIPTKLSNRYIPVSDRYKGYIYMEGSDDEKAFMSDTTQV--SSSEY 2549
Oy 1846 EELDINDIYVPGSKYKTLLEVLLEPGSNNTTASGKNTPSSTRNDIONDGSPSSKITDNE 1905
Db 2550 EELDINDIYVPGSKYKTLLEVLLEPGSNNTTASGKNTPSSTRNDIONDGSPSSKITDNE 2609
Oy 1906 WNOLKKEPISMLONOPNDVNDYSGNSSTNTITTTSRHVDNNTNTTMSRDNMEENL 1965
Db 2610 WNTLKDEFISMLONONEBTEPN-----MLGYVNDNNTHTPTSSRHNEKBP 2654
Oy 1966 LLPRIHGNILXSGEESYNNVNMVNSMNDIPINRDNNYSGIDLINDLSGSKPDIYDEV 2025
Db 2655 FIMSIHNRDLYSGEESYNNVNMV--NDIPISARNGNSGIDLINDLSNSKR--VDIYDEL 2711
Oy 2026 LKREKNELFGEENTKR--TSTONVAKTNSDPIHNOLEFHWLDRHRDMCKEY--KNKE 2081
Db 2712 LKREKNELFGEENTKR--TSTONVAKTNSDPIHNOLEFHWLDRHRDMCKEY--KNKE 2771
Oy 2082 DIINKLKEENK-----NINNSGKTYSNDKPS-----HN--HVLNTDIVSIQ 2122

```

```

Db 2772 ELDLKLKEENKNNNSGINSNTPTSPDIPSGQSDIPEDNNHISDIPVLTVDVSIQ 2831
Oy 2123 IDMONPKTNEITMADTNOQKSTMDTJLIDLEK--YNDPIYDFEDDITIHVDUYEKSMM 2181
Db 2832 IHMONPKTNEITMADTNOQKSTMDTJLIDLEK--YNDPIYDFEDDITIHVDUYEKSMM 2181
Oy 2182 DDIVVDH--NNVTSNNMDVPKTHIEMNIVNNKKEIFEPEEYPISDIWN 2228
Db 2884 -----DHDSTIVDTNADDEPSKVOIEMD--VNTK--LVKEKPIPLDMDI 2924

```

## RESULT 10

```

C71625
variant-specific surface protein 1 homolog PFB0010w - malaria parasite (Plasmodium fa
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: C71625
R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
: Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: C71625
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1711 <GAR>
A:Cross-references: GB:AE001366; GB:AE001362; MID:q3845070; PIDN:MAC71792.1; PID:q384
A:Experimental source: clone 307
A:Genetics:
A:Gene: PFB0010w

```

Query Match 27.3%; Score 3297.5; DB 2; Length 1711;

Best Local Similarity 35.6%; Pred. No. 3.9e-142; Index 741; Gaps 87;

Matches 832; Conservative 265; Mismatches 502; Indels 741; Gaps 87;

```

Oy 1 MATSGSGGTQDEDAKHLDFEFGQVHDEYHGAENYVSELSIASILGETAFTVKS 60
Db 1 MATSGSGGTQDEDAKHLDFEFGQVHDEYHGAENYVSELSIASILGETAFTVKS 60
Oy 61 MOTESKYTELLEANSKRNPK-----KDGKND-----VDRFSKEQAGYNNK 104
Db 61 MOTESKYTELLEANSKRNPK-----KDGKND-----VDRFSKEQAGYNNK 104
Oy 49 TYPMDKHP-----EGSTENPKLQYDYNTNWTYHGGQEPCTDIVERFSPTEGACQCKK 105
Db 49 TYPMDKHP-----EGSTENPKLQYDYNTNWTYHGGQEPCTDIVERFSPTEGACQCKK 105
Oy 105 MKCSNGMTCAPEERLHLCNKNFPMNSNDSSKAKHDLAEVCAAXEYSIKTHYPKD 164
Db 105 MKCSNGMTCAPEERLHLCNKNFPMNSNDSSKAKHDLAEVCAAXEYSIKTHYPKD 164
Oy 165 SKYSGSDFPMCTMLARSFADIGDIIRGRDYLGNKKRKGKQKETEREKLDEQKLEIFK 224
Db 165 SKYSGSDFPMCTMLARSFADIGDIIRGRDYLGNKKRKGKQKETEREKLDEQKLEIFK 224
Oy 166 ETNPDTKSQCLTVLARSFADIGDIIRGRDYLGNKKRKGKQKETEREKLDEQKLEIFK 224
Db 166 ETNPDTKSQCLTVLARSFADIGDIIRGRDYLGNKKRKGKQKETEREKLDEQKLEIFK 224
Oy 225 HDNLK-----DKAQKRYNGDEDPNFYKLRDWTANRETWGCMTCSKELDNSSYFRA 278
Db 225 HDNLK-----DKAQKRYNGDEDPNFYKLRDWTANRETWGCMTCSKELDNSSYFRA 278
Oy 222 YDELKNGKTNEEELQKRYRGDKDNDYOLREDWMDANRETWGCMTCSKELDNSSYFRA 278
Db 222 YDELKNGKTNEEELQKRYRGDKDNDYOLREDWMDANRETWGCMTCSKELDNSSYFRA 278
Oy 279 TCNPTGGSPSOTHNKRCGDKDKANAKPKRAGDGVITVPPYPPYQYLRWPEEWAEDF 338
Db 279 TCNPTGGSPSOTHNKRCGDKDKANAKPKRAGDGVITVPPYPPYQYLRWPEEWAEDF 338
Oy 280 TCG-RGELPYVTLKCC-----TAG-----VPPYPPYQYLRWPEEWAEDF 322
Db 280 TCG-RGELPYVTLKCC-----TAG-----VPPYPPYQYLRWPEEWAEDF 322
Oy 339 CRKKKLEMLEKQKDKSDERYCSRNGYDCEOTISRKGVRMGKCGTDFPAGSY 398
Db 339 CRKKKLEMLEKQKDKSDERYCSRNGYDCEOTISRKGVRMGKCGTDFPAGSY 398
Oy 323 CRKKKKIIPVNTKNCQRYVNGKE--KYDDRGYNDGDIRRQYIRLDTDCCKSLACKTF 381
Db 323 CRKKKKIIPVNTKNCQRYVNGKE--KYDDRGYNDGDIRRQYIRLDTDCCKSLACKTF 381
Oy 399 ENWIDNRQKQKOK--KYTEISDGGGR--KRAVGTTKYEGYKSYEKLKNDGYCTVD 456
Db 399 ENWIDNRQKQKOK--KYTEISDGGGR--KRAVGTTKYEGYKSYEKLKNDGYCTVD 456
Oy 382 AEWIDNRQKQKOK--KYTEISDGGGR--KRAVGTTKYEGYKSYEKLKNDGYCTVD 456
Db 382 AEWIDNRQKQKOK--KYTEISDGGGR--KRAVGTTKYEGYKSYEKLKNDGYCTVD 456
Oy 457 AFLGLLNNKACKDITGCKINFEVNSGGGVGSGGSGGSGASGCTNDEK--GFTYSEY 515
Db 457 AFLGLLNNKACKDITGCKINFEVNSGGGVGSGGSGGSGGSGASGCTNDEK--GFTYSEY 515
Oy 441 SFLOLSKERICK-----RIQVGEETANYG-----NFEENSFTSHTEY 480
Db 441 SFLOLSKERICK-----RIQVGEETANYG-----NFEENSFTSHTEY 480

```

OY	516	COCPDPCGVOKHGGONQWNERKT -VKMRMSKTYKPIPIGKNVLLLSLKLYVKDMHLLKKNM	574
Db	481	CDRCGLGVDCSSDNCRKKKPKPKSCDEBITDKETPEPTTTKIPKLTAEK - RKTGILTK -Y	537
OY	575	KEPCLTOWSSDGSVSVYVTGASGNSKEKELDYDEMCKYHNEVOKVWVGEEEDDEL	634
Db	538	KRCF- -KNSGN-----NGOJLTK-----WIC-HYE-----KNDKDC	567
OY	635	KGAGGLCLPLPKPKNKEVSEAKSONNHADIOKTFHDEFYVVAHMLKDSIMHRTKRLKS	694
Db	568	NGDINNCIOGDMKTSKNVYPIIS-----YSPFYGISIDIMLNESEIEM-ERLKS	616
OY	695	ISDKMTKCRNGCKKKCKDCFEKMYKOKETEMKPIKHOFKQ-----GIPGYFTTL	747
Db	617	INDKMLGCRGCGKNPECYKRWKKEKDDMDKKEFFRKOKDLKDIAOMDG--ELL	673
OY	748	ELIKLOFLKEDTEENTENSLOAEAEELKHIOKILLENMNLAVYNACTEOKTLMDXL	807
Db	674	EFLYLENIFLEMKRANNDPKV-----IEKFEILKEKME-----YQDLKTKRTITDPE	722
OY	808	LNHLENDATKC-----KDCPLPEEDKSRRGSADPSPDIFIPRE-----EKEDENED	855
Db	723	LEKELNEAKNCVEKNPNNECP--KOKAPGDGAAPSP--PREDITHDGEHSSEDEE	776
OY	856	DDDE-----EYDDEETAKETEGSATDITTSIDYCPYGVKLR	895
Db	777	EEEEEDDOPPAEGTEEOGESEKSEKVEVEODETPOKDIETKVPPTTTPVADCDTV-KTALA	835
OY	896	DNESLODACSLKY-GGNNSRLRGMCVCPSPGEPSTSSONKACIYCPPRRL--YIKKIY	951
Db	836	DTGSLNMAACSLKYTTGKN--YGMRCIAPSG---TTSKDDAICVPPPTOELCLYILEKS	890
OY	952	DMATKTESPOASGSEASSTSGSTTPPDSKEALLKAFUESAIIETFFLMHYKKEEKRAVAQ	1011
Db	891	D-----TKOGLREAFIKTAAGETIYLLMQYKXKDKQN---922	
OY	1012	EGAGHGLPRVEGSPRY--DPEKLEKGIKIPDGLFOMYTTIGDYRIDLIFSG---SNDT	106
Db	923	-----ETASTELDIDDPOLNGELPEIDPFKROMFYETGEDYND-LELGRYIGNDL	971
OY	1066	TSVSKDTPSSNDMLKINVLASSTEOGERKNMKYKEINFRCSTERSAPMLVSHPT	112
Db	972	DKV-----NNNITAV-----FONGCHITNGOKTDRQ-----QE	100
OY	1126	WMENNGKXIYHMGVCAITLSKDKIAKGYEKKROKLEPNENLMEANKKPKPOYOYTNVKL	118
Db	1001	FMGTGKXIOIKMGMLCALQ-----EAGGKTLTETYNYSVTF	103
OY	1186	DENGSTSPRTTOTOASDNPPTTLTHVVKRPTRYRWFEEWGESFCRRKRLKOIVDCK	124
Db	1038	NGHL-----TGTKLINEPASRPSFLRMWTEGDOFCERRITQLOILKERCM	108
OY	1246	V--ENGDVGRCSGDGEACDSISTHDYSTVSPFNOGCGKCSYRKRIEKKLIEFHQSN	130
Db	1083	VYQINGDGKDKDKK-----CTEACTYKKEIYIMWQNYKQNO	112
OY	1304	AYGOK--TDATRNNGNTFDEKCEKTELETPDAKFLER-LKNGCHKYNKEGGDDIFEK	136
Db	1123	RYTEVKXGSPRKESDVKESY-----AHQYLAKILKNIITCS--GTDI---	116
OY	1362	DSKTFQHTGYCGPCPKFRYNCQNGCVSGLNGCNDGKSIDAKEIAKMSSTTDVVMRY	1422
Db	1165	AYC-----NCMEG-----TSTPD-----117	
OY	1422	SDNPNTFEGDLDLADACOHANIFIGIKKDWKVCYGVGDICQDTNINERTDQEIQR	148
Db	1178	SSN-----118	
OY	1482	ALFKRWENFLEDYNNKINDKISHCIRKKGEGSCINGCESKSCLEKWIIEKIAEWENIKK	154
Db	1181	-----NDNIPESLK-----118	
OY	1542	RPNDQYENKDDPDYNVKSIIEELIPKIAVYNDODNTIKLCVFENSKGCTLLISMTQNNKEN	160

Query Match 26.4%, Score 3192, DB 2, Length 2706;  
Best Local Similarity 30.9%, Pred. No. 4,4e-137;  
Matches 885; Conservative 309; Mismatches 622; Indels 1052; Gaps 104;













1731 ----- 1730  
Db 2473 RRRKDCFPNDLKNIQENEVDNSLLEKIQHVAKNEGIDILKKLPNDONAASEICDAMKY 2532  
Qy 1731 ----- 1730  
Db 2533 SFADGDIIRGRSKIDPTNNKIEKELOKIFQIODNANSLKMLPELREKMMANRKE 2592  
Qy 1731 ----- 1737  
Db 2593 VWNAMTCVAPNDADHLKKKNNPGNKSQIIASOTEQTKKSHDEPPDYDIPEYRFLQE 2652  
Qy 1738 ----- 1744  
Db 2653 WSEYCKALKERNDKMDCKSCIKSGATCEKEDEKCKECCNDCKEYKNIVDKWQSEF 2712  
Qy 1745 ----- 1744  
Db 2713 DQGNOLYKLYTQDRHGRPSARARNPSTIEFQKLEDSGNDPYSAKYLDISTHCTDYKFS 2772  
Qy 1745 ----- 1744  
Db 2773 ETDRESNYAFSPYKDYKENCCKVNTPTSNNDKSPSLGSPFLPKPKMKFYPKIG 2832  
Qy 1745 ----- 1758  
Db 2833 IGVLPFTNMVADPTIHETVAKTFPNNAVPOPHINPDKTDVAPPTKNTLNEVLPSPAIYVG 2892  
Qy 1759 VALAGSTAFLELKKTKRASYGNLFQILQIPKSDYDITPLKSNRYIPVSDRYKKG--TY 1817  
Db 2893 IALAGSTAFLELKKTKRHPV-DLFSVINIRKSDYDITPLKSPNRYIPVSDRYKKGNGTF 2951  
Qy 1818 IYMGDSDEDKYA-FMSDTTDTVTSESESEYELINDIYVPGSPKXYTLEVVLEP---S 1872  
Db 2952 TLKEIVGTDSGYTDHYSIDT--SSSESEYELINDIYHVLNKTLEVVLEPSSGKLS 3009  
Qy 1873 GNNTTASGKNTPSDTRN 1889  
Db 3010 GNTIPTSGKNTPSDTRN 3026  
RESULT 15  
T14602  
variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragments)  
C:Species: Plasmodium falciparum  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jun-2000  
C:Accession: T14602  
R:Voss, T.S.; Felger, I.; Weise, N.; Beck, H.P.  
submitted to the EMBL Data Library, February 1998  
A:Description: Identification of a conserved 5' flanking region of Plasmodium falciparum  
A:Reference number: Z18158  
A:Accession: T14602  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2135 <VOS>  
A:Cross-references: EMBL:AF050740; NID:g2944094; PID:g2944095; PIDN:AAC05220.1  
C:Genetics:  
A:Gene: varph17

Query Match 20.8%; Score 2518; DB 2; Length 2135;  
Best Local Similarity 29.4%; Pred. No. 1.3e-106;  
Matches 676; Conservative 273; Mismatches 523; Indels 824; Gaps 82;

Qy 1 MATSG--SGGTODEDAHVLDPEFGQVHDEVGAKNVSELKSLAS-----ILGE 53  
Db 1 MAPEDDPGGGGTKDDAKHMDRIGTIOKOVHNAAEFRNLQGRLSATINIKLIG- 59  
Qy 54 TAFVYSKQTESKYTEL---EANSKRNPCKKDGKNDVDFSVKEQAGYDNKTKKCSN 109  
Db 60 ---NIETCSLEYKYYKHNGGADVSDKRYPKCKELSR-KYVDRFSKIGGCGCTNEKMRBG 115

Qy 110 GMTCAFRRLHLCNKNFPNMSN--DSSKAKHDLAEYCMAAKYEGES--KTHYPKYDSKY 167  
Db 116 IGACAPYRRLHLCNHNLETIFNNTESNNAKNLNVDCMAAKYBGDS:KTYTTHQHKY 175  
Qy 168 PESDFPMCTMLARSPADIGDIIRGDIYLGKKNKQNGKETEREKLQO:TLKEIFKRIHN 227  
Db 176 --DSQLCTVLARSPADIGDIYRGKDLFYGMTQKE-----KREDLEN LKIDIFKIHSG 228  
Qy 228 LKDKAOKRYNGDEDPNTYKLRDMMWTANRETVWGCAMTCSKELDNSTPRATCNDTGGP 287  
Db 229 LIG-GVKARYK--DTNTEYELREDMMWTANRETVWMAITCDA---HGTYFRATCGDNISP 282  
Qy 288 SOTHNKRCDDKCANNAKPRAGDGYTTPYFPYVOYLWPEEMEDRCRKKKLE 347  
Db 283 SMAKNNCRQCKKDGPRDDO-----VPTFYVPOFLWPEEMEDRCRLKHKLK 332  
Qy 348 MLEKCRGKDRSDERYRSGRNGYDCEQTIISRKGVRMKSGCTDCFPACGSYENMIDNOK 407  
Db 333 DALQCRKREKHKAKKLYDLDNRYDCEQTAGSKHDFEEDDCGCGQYSCARFVMTIDNOK 392  
Qy 408 QPDKO-KKTYTEISDGGGKRAVGG--TTYEGYEKFEYKLNKDGICYDAFLGLLN 464  
Db 393 EFLKOKKKYGTETIS-LKSRKKRDAGISTKYVDGEKKFEYELKSE-YRTVGEFLGLLN 450  
Qy 465 EKACKDIDG--GKINFEVNSGGGVVGGSGTSGAGTNDENKGTLYRSFYCOPCPDG 522  
Db 451 EYTCKEVDDKEGKTIDFTVNSG-----SASG-DDVVK-TYRTYTCACAPWC 496  
Qy 523 GV-QHKGNGQWERK-----TKVKMRMSKLYRPINKMVL:KSLKVNDMIL 570  
Db 497 GAEOGRNCGVMKAKNDRCSPGNDYTKYK-----KEPLTJGDKTKSEIY-- 542  
Qy 571 KKNMKEFLT--QNSDGSVGV-----VTTGASGNSGKKELYDEKCY-- 613  
Db 543 -ERYRKFCKNKNKANREGGVGSENGAASNDATTVGCGGNSJSLCEKTKCYX 601  
Qy 614 ---KNEVQKVN--VQGEVEEDDELKGAAGLICLPNKKKNEVEASAQNNAIDQTF 668  
Db 602 KKENNDGRKDNFCVQAGMO-----NSKKDKV-----RSY 632  
Qy 669 HDEFFYVVAHMLKDSIHMTKRLKSCIS--DGKTKKCRNGCKKCDCEKVVYKOKETEM 726  
Db 633 NAFPDWYHDLIDISIKRKNENGK-CINKDNKT--CIKGSCKDCFLKVVQOKREKMK 689  
Qy 727 PIKDFKTOEGIEGY--FT---TLELIL-KLOFLK--EDTEENTENSIDAEAEEL 776  
Db 690 LILHEHNTQGGFDKGEHQRGLFTHDVVLNVLDDKKEKLIKIEGTYGIR-----EET 740  
Qy 777 KHLQKIL-KLENENLAVYVNAAGTEBOKTLMKILNHELNDATYCKKCILPEE----- 826  
Db 741 KHKEMLDKEREKDAAGTGAAGTGPKNIMDKLIEHELQAKCKCKQCPPOSLGRSLNPH 800  
Qy 827 ----- 830  
Db 801 VYDDGSPKKRDKKRTNPCTSDTTTEYAVLAGVNAQFOGEVRAKMLJASRKNGETKSLE 860  
Qy 831 -----GRSA----- 834  
Db 861 GDIKKQPFKNGRSGSELNGDICKIDNKYSNDIRGSTAGPCTGKDGJNBRFNAGTKWEGD 920  
Qy 835 ----- 834  
Db 921 NEVSATHKLNLYIPRROHCTSNLEKLDLVTSKSNVNSDFLGDVLLAANNEAQRTRDH 980  
Qy 835 ----- 834  
Db 981 FAHKKDDHGIACRSVYSPADLADIIIRGRDMMDKDDGAQKMDIKKTFGNLYESTLPK 1040  
Qy 835 -----DPGP-DIFTP----- 843  
Db 1041 GKYYDDORTPOYKQLRDWMFEANRDQYKAWVEKDIKCEDEP1PVDYIIPQRLMMWT 1100  
Qy 844 ----- 843

Db 1101 EWAEWCKVQSQBYDELKCKGCKINGKVOGCTSGSDCTPCAACCTYGQIKIPWEDQ 1160  
QY 844 ----- 843  
Db 1161 WNNMLLOTLLYQOAEETARYGTRVSGDVGDKDPVQOFLLELOKONSGKTYNTAAQ 1220  
QY 844 -----RPEE----- 847  
Db 1221 YIHGEARVGECEVQKPCNTNGNQDKYFRREKPKDHDEACKCKDRPOQSAGGAGARSLP 1280  
QY 848 -----KEDDENEDDED-----EVRDEBETAKETEGS--ATD 878  
Db 1281 SPRPVDSDDDHSSDEDEDEDEDGGAEDENDEPASEEYKDDTEDEVEETAVSOPAAP 1340  
QY 879 TTTSLDY---CPIYGKVLTKNESLQDACSLKYGNNRSLGRCV----- 920  
Db 1341 TTTTPGYTPACEIYKDLFEKPKNTFKEACTOKYGGNNRSLGKCTPTSGDKAATRGSGD 1400  
QY 921 -----TPSGEPTTSSDKNGALCVPPRRRLYIKK 949  
Db 1401 TTKQNDSESGSESHQRAKHTSDASEKSAKSGEPT-----GSGICIPRRRKLTVYG 1455  
QY 950 IVDMATKTESQASGSEASSTSGST---TPDSKEALLKAFESAAIETFFLWHRKKEE 1005  
Db 1456 LTKWA-----EIOSQSQALSGQTPAGTPSOADPLAFAVESAAVETFFLMDRYKKL 1509  
QY 1006 KKAAYQAGAGHLP-RVEEGS--PEYDEPKLESGKITPDGFLROMFTLGDYRDLFSGS 1062  
Db 1510 NAPOGSSLLGGAPLQLNGAIGSEETPELSKGTIPPDFLRLMFTYLLGDYRDLGVGK 1569  
QY 1063 NDTTSVKDTPSSSNDLKNIVLLASGSTEOREKMKYKEI-KNFRKCTERSAPNLVS 1121  
Db 1570 QDVTKALE---ASGDN-----KSSKNPMQELSSKIEILKN-----GCTPPPTPVY 1612  
QY 1122 H-----POTWENNGKYYTHGAVCALTSKDKTAKGVKKPKIENPENLMDANKPK 1174  
Db 1613 HSPSSGTTTSSWMKTNQOHIKMGICALTYKE---SGEKKIEQVKATDN--TDLFEKIK 1666  
QY 1175 PROQYNTNVKLDENSGTSPRTTQOASSDNP--TTLTHFVKRPTVFRWPEMGESFCRE 1232  
Db 1667 DKYSYDKVEIKETENETBAKPO---DGLTPOTLLSNEFKRPFRYLEWQONFCKT 1722  
QY 1233 RKRLKQIKYDCK-VENGDVGRCSGDEACDSISTHYSTVPFNCPCGCKHCSYRKMI 1291  
Db 1723 RKRLKDIIECRNSDNPBGHGYCGDGHCHNELKN-NMFGIDCRDCKECCRKKKV 1781  
QY 1292 ERKIEFHKQSNAYGOOKTDATRNNGNTF---DKPECKTLETWPDAAKFLERLKNPC 1346  
Db 1782 HKKEDEFHNOKNKYGEHEKLT--NGDNYSGGDNTNFCCQIKEKKTAEDELKALRH--C 1837  
QY 1347 KT-----NKEYGGDDIDPEKDKSTFQHTPEYCGPCPKFTNCONGNG-----VSGLNGN 1395  
Db 1838 KNSDDTJKSEDEKKNINFEKPTNPTSTYKACAPYGVTCNRGTGRGCRPKIINEKN 1897  
QY 1396 CDGDKSIDAKEIAKMSSTVDVVARVSDNDNTNTEGDDLKDACOHANIFKIRKDVWCKG 1455  
Db 1898 IEGEQ-----IDINILIDDGATNDTD-NELHEKCKEYGLYTNLKKQEWKCO 1942  
QY 1456 YVCGVDICE-OTNIN-ERTDGEKEYIQIRALEKRWVENFLEDYKNINDKISHCIRKGECSK 1513  
Db 1943 HKNSYKCELOKPLNSEYD--ERIPKILFERMIIDIFIQYNNKSKEIRTC--TNDVNS 1998  
QY 1514 CINCCEKNSKCLEWIKETIAEMENIKKRPNDOYENKDOF-DYVAKSILEE-LIPKTA-- 1569  
Db 1999 CKECKGNCDCVEEMLKKSAMEKIKDYNOHPODVERIASRTKSFEBGTFPSYAKK 2058  
QY 1570 ---VVNDODNVIKL--CVFENSKGCTLISNTQNNKENDAITDCKLGVAKKNCPGKPSG 1624  
Db 2059 AQEYVEKREDEKIMGCTGSNTDG-----QTEBENGDFITNLDRLOTKIYQYQOTOHMK 2113  
QY 1625 EKQSDCKEPPPLPDEE 1640  
::: | | | | |

Db 2114 TORNSC--HPPPNDEE 2128

Search completed: May 1, 2001, 13:06:18  
Job time: 251 sec



THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2001, 13:02:07 ; Search time 52.89 seconds

(without alignments) 2408.011 Million cell updates/sec

File: US-09-508-967-1

Perfect score: 12100

Sequence: 1 MATSGSGCGTQDDAKHVLD.....VNKKKEFEDEYPISDIWN1 2228

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.GeneSeq\_0401.\*  
1: /SID56/gcgdata/geneSeq/AA1980.DAT.\*  
2: /SID56/gcgdata/geneSeq/AA1981.DAT.\*  
3: /SID56/gcgdata/geneSeq/AA1982.DAT.\*  
4: /SID56/gcgdata/geneSeq/AA1983.DAT.\*  
5: /SID56/gcgdata/geneSeq/AA1984.DAT.\*  
6: /SID56/gcgdata/geneSeq/AA1985.DAT.\*  
7: /SID56/gcgdata/geneSeq/AA1986.DAT.\*  
8: /SID56/gcgdata/geneSeq/AA1987.DAT.\*  
9: /SID56/gcgdata/geneSeq/AA1988.DAT.\*  
10: /SID56/gcgdata/geneSeq/AA1989.DAT.\*  
11: /SID56/gcgdata/geneSeq/AA1990.DAT.\*  
12: /SID56/gcgdata/geneSeq/AA1991.DAT.\*  
13: /SID56/gcgdata/geneSeq/AA1992.DAT.\*  
14: /SID56/gcgdata/geneSeq/AA1993.DAT.\*  
15: /SID56/gcgdata/geneSeq/AA1994.DAT.\*  
16: /SID56/gcgdata/geneSeq/AA1995.DAT.\*  
17: /SID56/gcgdata/geneSeq/AA1996.DAT.\*  
18: /SID56/gcgdata/geneSeq/AA1997.DAT.\*  
19: /SID56/gcgdata/geneSeq/AA1998.DAT.\*  
20: /SID56/gcgdata/geneSeq/AA1999.DAT.\*  
21: /SID56/gcgdata/geneSeq/AA2000.DAT.\*  
22: /SID56/gcgdata/geneSeq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12100	100.0	2228	20 W93944	P. falciparum PfEM
2	5074	41.9	2197	21 B18352	Plasmodium falcipa
3	4817	39.8	2182	18 W22476	Plasmodium var-1.
4	4817	39.8	2182	21 W77906	Plasmodium var-1 p
5	4802	39.7	3060	18 W22475	Plasmodium var-7 p
6	4802	39.7	3060	21 W77905	Plasmodium var-7 p
7	3928.5	32.5	2913	17 W00384	Plasmodium falcipa
8	3197.5	26.4	2703	16 R70236	P. falciparum Proj3
9	3197.5	26.4	2710	18 W22482	Plasmodium Proj3
10	3197.5	26.4	2710	21 W77904	P. falciparum Proj3
11	2929.5	24.2	1726	17 W00385	Truncated Plasmodi

12	2497	20.6	1700	21 B18144	Plasmodium falcipa
13	1385	11.4	700	16 R70235	P. falciparum EBL-
14	1385	11.4	700	16 W22481	Plasmodium ebl-2.
15	1385	11.4	700	21 W77903	P. falciparum ebl-
16	1319.5	10.9	431	21 B18350	Plasmodium falcipa
17	782	6.5	440	21 B18146	Plasmodium falcipa
18	702	5.8	921	21 W22480	Plasmodium E31a.
19	702	5.8	921	21 W77902	P. falciparum ebl-
20	697	5.8	793	16 R70234	P. falciparum E31a
21	428.5	3.5	1604	16 R70105	TNF-R-EBA 175 fusi
22	425	3.5	1786	14 R41043	CD4-EB4175 fusion
23	418	3.5	1435	16 R70232	P. falciparum SAMP
24	418	3.5	1435	16 W22477	Silic acid bindin
25	418	3.5	1435	21 W77900	P. falciparum SAMP
26	406	3.4	248	21 B18151	Plasmodium falcipa
27	379.5	3.1	362	21 W77912	Plasmodium DBL gen
28	360.5	3.0	411	21 W77913	Plasmodium DBL gen
29	345.5	2.9	749	16 R70233	P. falciparum EBL-
30	345.5	2.9	749	18 W22479	Plasmodium ebl-1.
31	345.5	2.9	749	21 W77901	P. falciparum ebl-
32	325.5	2.7	1245	16 R70106	TNF-R-Pl. viVax Du
33	321	2.7	2441	21 B18161	Plasmodium falcipa
34	315	2.6	1028	14 R41044	Plasmodium vivax D
35	315	2.6	1115	12 R13457	Duffy receptor. P
36	315	2.6	1115	18 W22478	Duffy antigen bind
37	315	2.6	1115	21 W77899	P. vivax DABP bind
38	314	2.6	1061	16 R70231	P. vivax DABP. Pl
39	311.5	2.6	1663	15 R46608	Plasmodium falcipa
40	310.5	2.6	1979	21 B18171	Plasmodium falcipa
41	287	2.4	1308	21 B18167	Plasmodium falcipa
42	287	2.4	1588	15 R46605	Malarial PfEMP3 ep
43	272.5	2.3	2907	21 W57452	Human transcriptio
44	271	2.2	2485	21 B18172	Plasmodium falcipa
45	265.5	2.2	1639	19 W54145	P. falciparum synt

#### ALIGNMENTS

RESULT 1	
ID W93944	W93944 standard; protein: 2228 AA.
XX	
AC W93944:	
XX	
DT 30-JUN-1999	(first entry)
XX	
DE	P. falciparum PfEMP1 protein.
XX	
KW	Erythrocyte membrane protein: EMP, PfEMP1; malaria; antioocclusiona;
KW	glycosaminoglycan-like moiety; antiaggregational; antimalarial;
KW	antigen receptor; infected erythrocyte; rosette formation; blood cell;
KW	capillary occlusion; cerebral malaria; treatment; vaccine; detection;
KW	medicament; parasite; diagnosis; drug screening.
XX	
OS	Plasmodium falciparum.
XX	
PN	W09915557-A1.
XX	
PD	01-APR-1999.
XX	
PF	18-SEP-1998; 98MO-SE01675.
XX	
PR	19-SEP-1997; 97SE-0003386.
XX	
PA	(KARO-) KAROLINSKA INNOVATIONS AB.
XX	
PI	Barragan A, Carlson J, Fernandez V, Qijun C, Whilgren M;
XX	
DR	WPI; 1999-254692/21.
XX	
PT	New isolated malaria polypeptides
XX	

PS Claim 4; Page 67-74; 80pp; English.

XX This invention describes a novel Plasmodium falciparum erythrocyte  
CC membrane protein (EMP), PfEMP1, which is capable of binding to a  
CC carbohydrate which exhibits at least one negatively charged  
CC glycosamino-glycan (GAG)-like moiety and has antiaggregational,  
CC anticoagulant and antimalarial activity. The carbohydrates of the  
CC invention are capable of acting as receptors for malaria antigens  
CC present on the surfaces of malaria infected erythrocytes, by binding  
CC to these antigens the carbohydrates prevent rosette formation by the  
CC blood cells, this prevents occlusion of capillaries as is seen in  
CC cerebral malaria caused by Plasmodium falciparum. The products of the  
CC invention can be used to treat malaria or to vaccinate against it, or  
CC used to design a model to identify compounds that bind to PfEMP1. The  
CC carbohydrates, polyphosphates and antibodies of the invention can be used  
CC as a medicament for dissolving the rosettes formed by erythrocytes  
CC infected by a malaria parasite. The products can also be used for  
CC detection, diagnosis and drug screening.

Sequence 2228 AA;

Query Match	100.0%;	Score 12100;	DB 20;	Length 2228;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2228; Conservative	0;	Mismatches	0;	Gaps 0;

QY	1	MATSGSGGOTDEPAKIVLDFEGOKVHDEVHGEAKNVVSLKOSLSLASLGETAFTVKS	60
QY	1	matsgsgsgotgedakihvldetfgkvhdevhgeaknvvslslasligetatlvs	60
QY	61	MOTSXTEELI EANSKRNPPCKKODKGNDVRFVSKEDAGVDNKKMCSNMTCAPRRH	120
Db	61	mtcsxteelleanskrnpckkodygkndvrfvskedaqydmkmcsmntcaprrh	120
QY	121	LCNNRFPNMNSNDSSKAKHDLAEVCMAAAVEGESITHPKYDSKPYGSDPFCMTLAR	180
Db	121	lcnkrfpmnsndsskakhdlaevcmaaelegesithpkyskpygsdpfmcmtlar	180
QY	181	SFADIGDILRGRLYLGNKKKKKONGKETEREBKLEQKLEIKELIHDLKDKAOKRYNGD	240
Db	181	sfaDIGdlIgrdrlylgnkkkkqngketerelEqkLeIkElIhdLkDkEaQkryngd	240
QY	241	EDPNFYKLRBDMWPNANGETWAGATCSKELEDNSSYPATONDPDGGPSQTHNKCRCDDK	300
Db	241	edpnlfykIredwPnanetwagatcskeleDnssyIratnddgypsgqthnkcrcddk	300
QY	301	GANGKRPKAGGDVTVIPTYFDVYPOYLRFMEAEADFCRRKKKKLENLKOCRGDKSD	360
Db	301	ganngkpkagdgdtvIptyfdyppoylrfmeaeadfcrrkkkklenlEkgcrgdksd	360
QY	361	EYRYCSNRNGYDEOJTISRKGVNRMGKCTDCCFPACSGSYENWIDNQRKOFDOKRYTKEIS	420
Db	361	eyrycsnrydceqtlstarkgvnmrgkctdcffpacsgyeniwdnqrkOfdkqtkykeIs	420
QY	421	DGGRRKRAVGCGTXXKYGSEKSEFEKELAKNGYGVYDAFLGLLNNKRAKRIITDGGKINRK	480
Db	421	dggrrkrravaggtxxkygsefEkelakngygvldaflglInnekackdItDggkInrk	480
QY	481	EVNNGGCVVGGSGGTSGASGTNDENKGTFRYRSBYCOPPCDGCVOHKGNGQMRKTRKVK	540
Db	481	evnnggcvvvgsggtsgaagtdendenkgtfryeeycgppcdgcgvqhkgngqwertkIvk	540
QY	541	MRMSKLYKPINGKMWLLIKSLKVVKDDMIILKKNMKEBCLTQNSSDGSVGSVYTTGASGN	600
Db	541	mrwsklykpingkmvllIkslkvkddmllIknmkecltqnsdsgsvgsyvttgsgn	600
QY	601	SEKELYDENKCYKHNVEYQVNVNOGVEEBDDDELKAGGLCIIIPNPKNKEVSEAKSQNN	660
Db	601	sekelydenkcykhnveqvnvngveeeddelkagaglcIIPnPkNkeVseaksqnm	660
QY	661	HADIOKTFHDFEYVWVHMLKDSITHWTRKLYKSCISDGRMKRKRNCGNKKCOCFEKVVQ	720
Db	661	hadIqkTfhdfIyvvahmLkdsIhwtrkLkscIsdgkrmkrcngcnkcdcfekvvq	720

QY	721	KETEKPKRKHFKQOEGIPBQYFTTLELILKLOFLEKEDOTEENTKSLDAEAEALKHLQ	780
Db	721	ketekpkrkhdfkfcqegipegyfftlelailkqlfikedteenstensidaeeaelkhld	780
QY	781	KILKLENNNLAAVNAQEBOKTLMDXLNLNELNDARKCKDCPLPEEDKSRGSAAPSPI	840
Db	781	kilklemnnaavnaqeboktldmcklmlnelndaekckdcpjpeedksrgsadspspi	840
QY	841	FIPRPEEKEDENEDDEDEYRDEERAKETTESGATDPTTSLDVCPIGKVLATDNESL	900
Db	841	fiprpeekedeneddedeyrdeerakettesgatdptttsldvcplygkvltkdneal	900
QY	901	QDASCLATGKNNSLGKRCYTPSGEPTTSSDKKGAICVPRRRRLYIKKYDMATKTSF	960
Db	901	qdasclatygnmsrlgwrctvpsgepttsdkgalcvprrrrlyikkivdwacktesp	960
QY	961	QASGEASSTSGSTPPRDSKLELLKMFVEASALIEFELWHRKKEEKAANOAGACHGPR	1020
Db	961	qasgeaasstsgstcpdpskrealikalvesaaliefilwrykxeekkaaqagahgpr	1020
QY	1021	VEEGSPERDEPDKLEKEKIFDGLFQMFYLGADYDILTEGSDNTTSYKDPSPSSNNL	1080
Db	1021	veegspeydpedkakegipdgflqmfylgdyrdlilesgndttsvskdpsnnl	1080
QY	1081	KNIVLLASGSTEOEREMNKYKEIKENRKCSTERSAPNLVSHOPTWENNKGITHHGVC	1140
Db	1081	knivllasgstegerekmnykelkifkctersapnlvshpctwennkgylthgvc	1140
QY	1141	ALTSKDRAKAVEKKPOKINPENLMDENAKRKPRYOYOTWVKIKDENGSTSPRTTQNA	1200
Db	1141	altskdrakavekkpkrlenpenlweanrkpkpqaycnvklidengspsrttcqna	1200
QY	1201	SSDNTPTLTATFVRPRPYPFMFEEBEGSEFCREBRKRLKLOIKYDKCKENDVGRCSGDEA	1260
Db	1201	ssdntptltltfvrprpyfifwfeewgesfcerkrlklkikvdkcngdvgrcsgdga	1260
QY	1261	CDSISTHDYSTVPSFNCPGCGKHCSSYRKWIERKKIEFHKOSNAYGOQKDTAFTNNGTF	1320
Db	1261	cdsisthdystvpsfncpgcgkhcssyrkwierkkiefhkgsnaygqktdafrnngtft	1320
QY	1321	DKECKTLETPPDAKRLERLAKNGPCKTNKEVGGDDIDENKSKPTQOHNTEVCGPCPKKT	1380
Db	1321	dkecktletpdpdaakrlerlknpgcktnkevgddidfeksktqtqheycgpcpkkt	1380
QY	1381	NCONGNCVSLANCGGDSIDAKELAKKRSSTTQVWVRVSDNPTNREGDLDKDACQ	1440
Db	1381	ncongncvslngncgdgdsidakelakmrstctvwmrvsdntntlegddlkdaeqh	1440
QY	1441	ANIFKIRKIDWKCXYGVGDICEQUTININERTDKEYIOIRALFKRWENFLEDYNNKIND	1500
Db	1441	anifkirkidwkcxygvdydicequtinerdtgkeyioiralfkrwenfledynnkind	1500
QY	1501	KIHSICKGBESKCIINGCEKNSKCLEKWIIEKTAEMENIKKRPNOYENKKODPOVNVMSI	1560
Db	1501	kishickgbskcilingceknsclekwieklaewenikkrfndygenkdpdvnvksi	1560
QY	1561	LEELIPIKAAVNDONVITKICVFENSJGCLISNTONNKENADICOMLKILVAKKAKNGP	1620
Db	1561	leelipkavvndonvikcvfenskgtclisntcnkendaicdmllklyvakncpg	1620
QY	1621	KPSEKOSDCKEPPRLDEEDONPEENTLEPPKFCPTTQPEEKGGETCGNKKEBKDEK	1680
Db	1621	kpsgekdsdckepprlpedeqnpeentleppkfcpttcqppeekggetcgnkkeekdek	1680
QY	1681	KEESEEPKKEESGPAAEPPAPTAEBSEETETNPPPEPGTGPAAPESTPAPPTDTPPPLRP	1740
Db	1681	keeseepakeesgpaaeepaptaeasetetnfppepgtgpaapestpaptptdtppprlrp	1740
QY	1741	QADPEPSTLIQTITPRGVALALGASTAFLFKKKTATASGNLFOLIOIPIKSDYOIPIPLKS	1800
Db	1741	qadepfstliqtitprrgvalalgsatflfkktktasvgnlfqilqipksdyoiplpls	1800

OY	1801	SNRIIPVSDKRYKKRTIIVABGSDDEKRYAPMSUTTDVYSSESEYEELDINDIIVPGSRK	1860
Db	1801	sniylypvssdykkykylymegdsdedkyaifmsdtltvtseseeeyeeldindiyvgspk	1860
OY	1861	YKTLIEVLAEPSGNNTTASAGKNPSPDRNDIONDGIPSSKITPDNEMNNOLKKEPKSMNLON	1920
Db	1861	yktlieevleapsgnntcasgknpsdclrndiqnqipsskitcdnewnqlketismnlqn	1920
OY	1921	QPNDDVPNDYTSGNSSTVTMTNITTTTSRRHAVVDNNVTNTWMSRDMMEENLLPSIHNGNLVSGEE	1980
Db	1921	qpnvdpndytsgnsstvtmtttsrrhavvdnnvtntwmsrdmmeenlllpsihngnlvsgee	1980
OY	1961	KSYAVNMVNSMNDPIPINBNNVYSGIDLINDLSGGKRPIDITYEVLKRKENELFGEENTK	2040
Db	1961	kysavnmvnmndpipirndnvysgidlindsisggkrpidiyevlkrkenelfgeentk	2040
OY	2041	RSTQNVAKKTNSNPRIHNOLELFHKVLDRHRDMCKEKKNEDILINKLKEPMNKENINNSG	2100
Db	2041	rtsqnvaaktnsnprihnqlelfhkvladrhmckekknedilinklkewnkeninnsq	2100
OY	2101	KTYNSDNKPSSHNVHLNLTDVSIQIDMDNPKTKNETTNMDTNOQDKSTJMDTIIDDLEKYNDPY	2160
Db	2101	ktynsdnkpshnhvlnctvsiqidmdnprktneltmdntqdksstndtlldtlekynpy	2160
OY	2161	YYDFIEDDIHYHDVVEKSSMDIYVDHNNTVSNNMDVPTRKHIEMNIVNKKKEIFEEX	2220
Db	2161	yydfyeddiyhhdvveksmdidyvhnnvtsnnmdvpckmhlemivnnkkeifeex	2220
OY	2221	PISDIWN1 2228	
Db	2221	pisdwn1 2228	
RESULT	2		
ID	B18352		
XX	B18352 standard; Protein; 2197 AA.		
AC	B18352;		
DT	07-NOV-2000 (first entry)		
DE	Plasmodium falciparum chromosome 2 related protein SEQ ID NO:210.		
KX	Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;		
KW	antimalarial; malaria; protozoacide; infection; insecticide.		
OS	Plasmodium falciparum.		
PN	MO200025728-A2.		
PD	11-MAY-2000.		
PE	05-NOV-1999; 99MO-US26796.		
PR	05-NOV-1998; 98US-0107131.		
PA	(HOEFF/) HOFFMAN S.		
PA	(CARU/) CARUCCI D.		
PA	(GARD/) GARDNER M.		
PA	(VENT/) VENTER J C.		
PI	Hoffman S, Carucci D, Gardner M, Venter JC;		
DR	WPI: 2000-365347/31.		
PT	Proteins encoded by chromosome 2 of the human malarial parasite,		
PT	Plasmodium falciparum, useful as antimalarial vaccines and in the		
PS	Diagnosis of P.falciparum infection -		
CC	The present invention describes proteins and their fragments (I) encoded		
CC	by chromosome 2 of the human malarial parasite, Plasmodium falciparum.		

```
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II).  
CC (I) and (II) are useful for the development of vaccines against  
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with P. falciparum. Furthermore,  
CC (I) (especially when they are rifins or secreted or membrane proteins  
CC can aid the identification of drugs to treat or prevent P. falciparum  
CC infection, or they can be used to identify drug resistance in  
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. A70078 to A70287 and B18144 to B18352 represent nucleotide and  
CC protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.  
xx  
xx Sequence 2197 AA;  
SQ  
  
Query Match 41.9%, Score 5074; DB 21; Length 2197;  
Best Local Similarity 47.1%; Pred. No. 6.9e-296;  
Matches 1134; Conservative 277; Mismatches 599; Inlets 400; Gaps 87;  
  
OY 4 SGGSGGTODEDAKHYLDERGOKVHDEVEHGAKNYSLSGLSTIASISLG-ETAFIVSMQ 62  
DB 3 sgkggdqpdgesvkhmfdrigevegvkselvnyslegrkislapiigvesgsnelcn 62  
OY 63 -TESRYTELLEANSKRNPCKKDKNDRFSVKBOAGYDNKKMKCSNGMT-----C 113  
DB 63 lvgdylnpyvgnnsrrpcknklgltneerfedtlggctcnkklgneystkskgdcgac 122  
OY 114 APFRRLHCNKRFPMNMSNDSSKAHDLLAEYCAAKTEGESIKTHNFKYDSKITGSPDP 173  
DB 123 apyrllhlcsh---nlseidtsmtkhlllevcmaakyeagnsidthpqhqrtnedpsq 179  
OY 174 NCTMLARSPADIGDIRGDILYGKRRKKONKEERREKLBOCLKIEPKKIHDNLKDE- 232  
DB 180 icmltarstadigdvlrvgtldilfngskeke----krdelelnlktfgkllekikdreg 234  
OY 233 AKRRTNGEDDEPFNYLRREDMWTANETWGMATCSKELDNSSYPFPA.CNDTGOGPSQHN 292  
DB 235 acetr-gsdtnhyqlredwyanaratweatlc--dhgsydfiq:sgdketatarxd 291  
OY 293 KCRCDKDGANAAGRKPRAGDDVTIVPYFYFDVYPOYLRFMEEMADEP:IRKKKKILEBKQ 352  
DB 292 kercddengkkpg-----snadgyprtyfdypylrfeweadef:rkkkkkklelegq 345  
OY 353 CGGRKKSDEYR---YCSRNGYDCEDTISRKGVRMGKCTDGFPGCJSEYMINDQRKF 409  
DB 346 cr-----dykanlylcsognugycrktlykkgylvgelctncsvcr:lyesvlddqk.lef 399  
OY 410 DKOK-KYTRKEIDSG-----GRRKRVAG--GTTKYEGYEKSEYELKLNKGSTV 455  
DB 400 lkqkqkyeteisnssgsgsgsvkvrnttkkrgayetatnydgskkykelkeeseykv 459  
OY 456 DAFGLLNNEKACKDITD-GGRINFEVNSGGVVGSGGTSGASGTNDENKKTFFYRSE 514  
DB 460 ddflllmmedvcfkikekedftk-----pdksnmegfyise 502  
OY 515 YGOPPCDGVGHKGGNOMERKTYVKMMRSKIYRIINGKNVLKSLKVAVDMMLTKNM 574  
DB 503 ykpcpdcgvvk-k-dnwkydkydgcctr-gkllypasasagtp.lkl:sgeqkxetekl 560  
OY 575 KEFCFLONSDDSVGSVYTGT--ASCENSEKRELTYDEMCKCYKHIEVOKV-----NVQ 624  
DB 561 kfcc--dtngdcltnsarvaggaagsksksnkelyeeakcy--newqykdkddangee 615  
OY 625 GEVEDDELKAGAGLCILPNPKRNKEYSEASKSONHADIOKTFHDFEYVAVHMLKDSI 684
```



Db 616 dedeadvkvkaggjcllen---khesinnassnepegfgktfnfdyfwjgrflndsm 672  
 QY 685 HMRPRRLKSCISDGRKTMKCRGNCKKCCDFEKWYKXKETEKKPIKDHKTQE----- 736  
 Db 673 ywrgk-vnscimnpkrkrkcnecddcgclfwgkkeeuenlknkhtkqeafrnkren 731  
 QY 737 -----GIPGYYFTTLELILKLOFLKEDTEENTENSLSAEEAELKHLKIL-----K 784  
 Db 732 sgldmfsglmds-advvalrelaleleqfqlkdg-----ygdvkelkylkelldeekkk 784  
 QY 785 LENENNLAVNVAAGTEBQKTLMDKLNHELNDATKCDCLPEEDKSRGASDPEDIFIP- 843  
 Db 785 kqaeeavvvvadnqkktldkllqhegdanncl-----kthkeceetpkp-----pg 835  
 QY 844 -----RPEEK-----EDDENEDDEDEVRDDEETAK--ETTEG----- 874  
 Db 836 agpgpapeetcttlededeedeedeagadeveegeelvdctlegdetetveqpykdtre 895  
 QY 875 -----SATDTTSLDVCPIVGVKVLTKDNESLDACSLLKYG-GNNSRL-GMRCVTPSGE 925  
 Db 896 geeeaaktdttsldvcdtlnalt-mdnldacklkypgpketfrmkcvs-sge 953  
 QY 926 -----PTSSDKNGAICVPRRRRLYI---KQIVMATITRESQASGEASSTSG 972  
 Db 954 kavalagsagaicgskgdgaicvprrrrlyvggllklltsagtsespg-gsaeasrsad 1012  
 QY 973 -----STPPDSKEALIKAFVESAALETFFLHRYKEE---KKAVAQE-----G 1013  
 Db 1013 vsqngsgddtltt-----eslrkfwiecaalefflwnrykweaqqkaelqngll19 1067  
 QY 1014 AGHGLPRVEEGSPEDPEDKL-KEGKIIPDGLKQMFYLLGDRILFSGSNDTTSVSKDT 1072  
 Db 1068 tgaal--nlgsgdaspqlqksgclpdlflmflyclggyrdllvgvad----- 1116  
 QY 1073 PSSNDNKNVLLASGTSEDEREKMKYKE-INKRKCS---TERSRPNLYSHPQTWME 1128  
 Db 1117 --dknggnmlllnasgn-kdekqmeklqekleqllpfsqnetketrpgnvnndrqs1wd 1172  
 QY 1129 NGKRYIMHGMVCAITLSK--DKIAGVEKKPQKIEPENLMEANKRPPOYOYOTNVLD 1186  
 Db 1173 rlaehvnhgmvcaltkydddnlgkgyvkkpkqklienpeklnetckkpkdekyqyqta1e 1232  
 QY 1187 ENSGTSPTTQTOQASDNTPTTLTHFVKRPYTFWFEEMGESFCRERKKRLKQIRVDCKV 1246  
 Db 1233 desg-----ekrpsasasgklcflkrppyfrylgeengfckrtlemigklkede-y 1285  
 QY 1247 ENGIVRGSGSGEACDSTSTHDYGVSPFNCPGCGKHCSSRYKRIETKRTIEFHQSNAYG 1306  
 Db 1286 knq--grcsqgllkenelivldkeklfgollcplcarhcrfyykwlntkrtdelfnkqsna5 1343  
 QY 1307 QOK-----TDATRNNGNTFDEKFECKTLETPDAKFLERLKNGPCKTNKE---YGGDD 1356  
 Db 1344 eqkkkyeendsaqknq-----vcgltk--ddaaeflnrlknpgckneeseenkkaede 1395  
 QY 1357 IDEFKDSTFQHTREYCGPFPKFTNCGNGCVSG-LNGNCDGKSIDAKEIARMKSSYT 1415  
 Db 1396 idfkppdtfdkdnckpcseefklckenhncsgntgkcdgkttlaateienlknltk 1455  
 QY 1416 DVNVRVNDNNTNTEFGDDLKACQHANIFKIRKDYMKCGVYCGVDY---EQGININERT 1472  
 Db 1456 evtlmvdsdksatefcfdg1seckdkg1fkg1rkdwecgkvvgvdicnllkdh1qhes 1515  
 QY 1473 DGEYIOIRALFKRMVENFLDYKNKINDKISHCIKKGEGSCINGCEKNSKCLKMWIEKK 1522  
 Db 1516 d-kkylimkellkrtwleyfledynlkhk1shcnkngyskcl9g-----cvdkwvqk 1568  
 QY 1533 IAEWENIKKRPNDQYENKDDPDY--NVKSIIEELIPKIAVNDQDNVILKLVFENSKGCTL 1591  
 Db 1569 keeqkqjkerfneqykaktsdeyfnvksfletw1pk1avndqgnv1k1skfngscgsa 1628  
 QY 1592 ISNTQNNKENDADIDCMKLCGVKAKNCBGRKPSGKOSDCKE---PPPLPEED--QNPE 1645  
 Db 1629 saistngneadaldcm1klekkideckkrpgensgqcnctllhpldvgedep1eete 1688

QY 1646 ENTL--BPPKCPPTTQPPPEKGGGTGKNKEKKDEKKEESEBPAKESSGPAEAPTA 1703  
 Db 1699 enpyvgkqnpstcpvpeakkkkeegetc-----tpasapa--- 1723  
 QY 1704 ESEETETNFBPPPGTGPAPRSTPAPPTPDPPLRQADEPFDSTILQTTIPRGVALAL 1763  
 Db 1724 -----papapaspaprpap-----adeqfcllqtllqtp1g1a1a1 1759  
 QY 1764 GSIAFLPK-----KTKASVGNLQIILIPKSDVDIPRLKSSNR 1803  
 Db 1760 gsiaflflkvlvlycvvymylymcfcilymkktkhrv-dlfsvaln1pksdyd1p1ks1pnr 1818  
 QY 1804 YIPVSDRKGYTYIYMGDSDEDEKAYAFMSDPTDVT--SSESEYELINDIYVPSPKYK 1862  
 Db 1819 y1p1tsgkyrgkry1ylegdsqtds-gytdhysd1tssesesejeemtd1ndyrgspkyk 1877  
 QY 1863 TLEIVLLEPSSGNNTTASGKNTPSDTRNDIONDGIIPSSKITDNENQOLKEFISMLQNP 1922  
 Db 1878 tliev1lepgnnta-----sdtqnd1qndg1psnkfsadnewntclkdff1sm1qnp 1931  
 QY 1923 NDVPRNDTSGNSTRTNTITTSKRNVDNNTTYSRDNMEENLLPSTHDGNLYSGEYS 1982  
 Db 1932 kdvpndyksgd1pfnctq-----pn1lyfdkpeekpflts1hdn1llngeys 1978  
 QY 1983 YVNMN-VNSMNDIPINRDNVYSGIDLINDLSGKRPIDYDEVLKRENELEFETWTKR 2041  
 Db 1979 yvnmnschnmdd-pkyvsnnvysgidlinds1gnkhd1ydeylkktkenelfgtnvkh 2037  
 QY 2042 TSTQNVAKTNSDIPINQOLEFHKWLDRHDMCEKWKKNKEDILNKLEWKNENINNSGK 2101  
 Db 2038 tsihsvakntdsd1l1nq1n1f1c1d1h1r1cm1cek1wen1her1ak1keew--eneethsn 2095  
 QY 2102 TYNSDNKPSHNHVLNTVYSIOIDMDNPKTKNEILNMDTNDOKSMMDI1DOLK-YNDPY 2160  
 Db 2096 thpsds-----nklntdvs1q1lmdn1p1n1q1fn-----mdcl1led1dk1p1nepy 2142  
 QY 2161 YUDEVEDDIYHDVDEKSSMDIYVDH--NNVTSNNNDVPTKNHIEINYNKKEIFEE 2218  
 Db 2143 yudmdd-lyydrn-----dhdstvtelnamdypskvq1emd-vntk-1vke 2187  
 QY 2219 EYPISDIWN1 2228  
 Db 2188 kyp1adw1d1 2197  
 RESULT 3  
 W2476  
 ID W2476 standard; Protein; 2182 AA.  
 XX  
 AC W2476;  
 XX  
 DT 12-SEP-1997 (first entry)  
 XX  
 DE Plasmodium var-1.  
 XX  
 KW DBL gene family; SABP: stalic acid binding protein; vaccine; therapy; Duffy binding like gene; Duffy antigen binding protein; erythrocyte; KW DBP; merozoite; malaria; var-1; var-2; var-3; var-7; Immune response; Plasmodium.  
 KW  
 OS Plasmodium vivax.  
 OS Plasmodium falciptarum.  
 XX  
 FN W09640766-A2.  
 XX  
 PD 19-DEC-1996.  
 XX  
 PF 07-JUN-1996; 96WO-US09508.  
 XX  
 PR 07-JUN-1995; 95US-0487826.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.



Db 1418 dewkancvqvdicltlekkikngqgdckkyltmkellkrlwylefiednrlrkkiklctk 1477  
QY 1508 KGESEKCIINGCEKSNKCLEKWEIKKIAEMENIKKRENDQYENKQOPDVNKA5IIEELIPK 1567  
Db 1478 kegdckctky-----clekwvgektkewqkndtyleqyknd-----gnltlnfleg 1555  
QY 1568 IAAVNDQDNVILKLC-----VFENSKGCTLI5NTONKKNKDAIDCMKLKLVAKNCPGKPS 1623  
Db 1526 fgyrtfefnaklpcpdgldqfktscglnsdtnsgnmdvlcllnklqkkliseckeghs 1585  
QY 1624 GEGKSDC-----KEP-----PPLPDEEQNPEENTLPEPKCPPTTQPREKGBE--TCG 1671  
Db 1586 gqctcpncaslsqksetlvedvddyeqnp-ekvgeqpkcpdmkepkndeavgctg 1644  
QY 1672 NKEKKEKKEESEPAAKEESGPAEEPAE--TAESEETETN-PPEPGTGAAPAPSTPA 1728  
Db 1645 gdeek--kvedsvleqkeeaasapeesplltpaepkkeenvpkpp----- 1690  
QY 1729 PPRPDTPP-----LRPADEP-FDSTIIQTTFPGVALAGSIAFLFKKTKASVGN 1781  
Db 1691 -----pppkrrlkrtnvldhpavlpalmastlmwsglfaaflylkkkcksavgn 1744  
QY 1782 LFOILQTPKSYDTPTLKSSKRYTPYSDRKGYTYMEDSD-EDKYAPMSOTTDVTS 1840  
Db 1745 lfqlglpksqdydipclksnrylpyasdrhkyklylmeqdsqdekyalmsdtldts 1804  
QY 1841 SESYEELIDINDIYVPGSPKKTLEVLVEPSGNNTTASGKNTPSDPTKNDIQND---GIP 1897  
Db 1805 sesyeeldindiyvpspkyllelvleps-----krdcqndlnhdipsdip 1853  
QY 1898 SS-----KITDNEMNQLKEPISNMLQNOPNDVPNDYTSGNSSTNTNTTTSRHVNDNTN 1953  
Db 1854 nsdclppltdedewnglkkdfismnlqntqtepn-----lhdhvndnth 1898  
QY 1954 TTNSRDNMEEELLRLPSHDGDLVSGEESYVW-----NMVW-----SMNDIPT 1996  
Db 1899 pcmrthmdqgprfmslndrlntfsggeeyuamfngnplnlsdstmsdlsnhspsy 1958  
QY 1997 NRDNVNSGIDLINDLSGCKPDIYDEVLEKRENELEFGENTFKRTSTQN--VAKTNSD 2054  
Db 1959 ndkndlysgldlndalsgnh-idlydeamlkrkenelfgcthpknltsnrvqtsd 2017  
QY 2055 PIRHQLLEPRKWLDRHRDMCEKMNKEDILNKLKEEMNKENINNSKTYNSDNKPSHNHV 2114  
Db 2018 pltnqlnlfhkwlgrhdmcekwnherlpklkelw--enethsgdt--nsjpsgnhv 2073  
QY 2115 LNTDVSTQIDMDNKTKNETNTNMDTNDOKSMTMDTILDDLEKYNDPYDFEDDIYHDV 2174  
Db 2074 lntvtslqldndpklmefnmdtpokstmdtildlekynepyydyfkhk-lyydv 2132  
QY 2175 DVKSSMDDIYVDHNNVTSSNMMDVPTKMHIEMNIVNNKKEIFEDEEYPS 2223  
Db 2133 nddkasedhlnmdhknmdnnsdvplnvglemynln-qellqneypis 2180

RESULT 4  
ID Y77906 standard: Protein; 2182 AA.  
XX AC Y77906;  
XX DT 13-JUN-2000 (first entry)  
XX DE Plasmodium var-1 polypeptide.  
KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;  
KW DABP; Stalic Acid Binding Protein; SABP; malaria; vaccine; immunisation;  
KW protozoacide; var-1.  
XX OS Plasmodium sp.  
XX PN US5993827-A.

PD 30-NOV-1999.  
XX  
PF 07-JUN-1995; 95US-0487826.  
XX  
PR 10-SEP-1993; 93US-0119677.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Slim KL, Chitlins C, Peterson DS, Su X, Wellens TE, Miller LH;  
XX WPI. 2000-194198/17.  
DR N-PSDB; 298288.  
XX  
PT Isolated protein binding domains from Plasmodium vivax and Plasmodium  
PT falciparum erythrocyte binding proteins useful for vaccinating against  
PT malaria -  
XX  
PS Disclosure; Columns 129-140; 93pp; English.  
XX  
CC The invention relates to ebl-1 polypeptides that are encoded by the DBL  
CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially  
CC identical to the Duffy Antigen Binding Protein (DABP) and Stalic Acid  
CC Binding Protein (SABP), which are soluble proteins that appear in the  
CC culture supernatant after erythrocytes infected with malaria release  
CC merozoites. Immunochemical studies indicate that DABP and SABP are the  
CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy  
CC and stalic acid receptors on erythrocytes. The ebl-1 polypeptides may be  
CC used to vaccinate against malaria, especially caused by P. falciparum.  
CC Immunization with the polypeptide provides effective protection against  
CC malaria. The present sequence represents the var-1 polypeptide.  
XX  
SQ Sequence 2182 AA:

Query Match 39.88; Score 4817; DB 21; Length 2182;  
Best Local Similarity 45.94; Pred. No. 2,4e-282;  
Matches 1096; Conservative 298; Mismatches 607; Indels 388; Gaps 93;

QY 4 SGGSG-----GTOD-----EDAKHVLDPEFGOKYHDE--VHGEAKNYVELKSGSLAS-I 50  
Db 11 sgsssgkygkdkdtegylyvsdaklllrvgekyeevkrngdakkyllealngntnng 70  
QY 51 LGRTAFVTKSMQTESKYTELLEANSKRNPCKDKGNDVDRFS-----VKDQAG 99  
Db 71 sseasslecltkeyeryerngqgrhpcrkaaknedvnrfdcltggctynrlksqg 130  
QY 100 YDNKMKCSNMGTCAPRRRLHLCKNFPNMNSNDSAKAKHDLAEVCMAAKYBESIKTH 159  
Db 131 gdnkv-----gacapyrlhldcy--nlesldtstltklillevcmakkyegnslnth 181  
QY 160 YPKYDSKYPGSDPFMCIMLARSPFADIGDIIRGRDLVYG--NKKKONGKETEREKLOKL 217  
Db 182 ytcqqrtnedsaagqlcvtlatstadlgdlvrgklylgydnkeqg-----iklqgk 235  
QY 218 KEIFKKIH-DNLKDEAKQRYNGD-EDPNFYKLEEDMWTANRETVMGAMCSKELDNSSY 275  
Db 236 kdlfkklhdkvmktnsgqeryldakgydffqledewtsnretcvkallchpakeanyf 295  
QY 276 FRATCNDTGQGPQSTHNKCRCDKDGANAGPKAGDGVTVIPYEDYVPOYLMEBEWA 335  
Db 296 lktacn-vgky---tngqchc-----lpgd-----vplyfdyvpqylwfeewa 335  
QY 336 EDPCKRKKKKLEENLEKQCRGKDKSDERYRSRNGYDCEORTISRKGKVRMGKGCIDCFAC 395  
Db 336 edlcrkrrkkklleqlqkrcryeqn---lycsngydccltlykqklylvehcnscwvc 392  
QY 396 GSYENMIDNORQFQKQ-KYTKEISDGG-----RKRRVAGGTTRYE--GYEKSEFEKL 447  
Db 393 rmyetwldngkkelkqkrkyeteisgsgsgkpekrktrraarssssddngyeakfykkl 452  
QY 448 KNDGYTVDAFLGLNNKCAKDDITDGGKINFEVNSGGVGGSGGTSGAGSTNDENK 507  
Db 453 keyvgdvdfklklkneqclqkqpyg--nekadn-----vdfitnekv 495

Qy	508	CTFPRSEYCO	PRDCPCYQ	NGHGGNOME	RKTKVKKMRR	SKLTKPI	NGKMLLSK	LKVWDM	567					
Dh	496	klferteicp	pcwgcgle	-kysppw--	kvygdktcgsa	tktycdph	nltdipvl	yipkdsq	552					
Qy	568	MILKMKRECL	QNSSDS	GVSVYTTGAS	CGSSEKKE	LYDMCKY	KHNHGVK	VWQGV	627					
Dh	553	gnllkkykntc	-----	ekgapgygqylk	-----	wqcy	-----	580						
Qy	628	EEDDEKJGAG	GLCILE	PNPKNEVSEAK	-----	SONNHAD	IQKTEFDFY	VWAMKLDS	683					
Dh	581	-----dehr	-----	psssmmncwgc	gtdktitgkyc	-kxynvffw	dkwvhlmds	626						
Qy	684	IHMFKRLKSC	---	SDGKTMKRC	MCKNKKDC	FEKWKQKETE	KPIKHNKTEOG	I--	738					
Dh	627	vwkte-1sk	innmtngt	cnrmkckldc	gcfkwkkg	qgwmalktd	gtgktdlvq	685						
Qy	739	-----PEGYFT	LIELLK	-----	LQFKEDTE	ENTENS	LDABEELK	MLKLE	786					
Dh	686	qkqlivfspyg	-----	vidlylkggnll	gnlk-dvngdt	-----	ddlkhlkkl	-ld	729					
Qy	787	NENNLAVN	NGTBEOKT	LMDLHNL	ENDATKCK	-----	DCPLDEBK	SRRSRSDPDIET	842					
Dh	730	eedavavnl	ggkntcti	-dkllqke	geaqcqbke	eeckkaqges	rjyrseet	derct	788					
Qy	843	PRP-----	BEKEDDE	NDEDEDEVR	DEDETA	KETTESAT	YTTSIAD	-----	885					
Dh	789	qqpdsagv	eeeee	dddydded	dddvqeeegc	---eglvte	rtetveet	veeq	846					
Qy	886	-----CPVGC	VLTKDNE	LQDACSLKY	-CGNNSHL	GRVCYPSGE	PTTSSDKNGA	ICV	938					
Dh	847	egvpcpd	lvqk-1	feddksikea	cgilkypg	gkcpmckv	cvtpsg	statsqkgaicv	905					
Qy	939	PRPRRRRL	YIKKIYDM	ARKTESP	ASGSEAS	STS	SGSTTP	DSK---	995					
Dh	906	pprrrrlyvg	lsglswast	-----	ggdettevs	seatsps	seesekl	rtaltesa	aalet	959				
Qy	996	FFLMHRKKEE	KAAVA--	OEGAGHL	PRVEECS	PEYDEDLKE	-GKPROGL	KOMTYGLD	1053					
Dh	960	flfkhkykeek	kpptacdg	dgglvgs	lpepspge	odqqlq	gtvylpr	pfllmqmtytlad	10139					
Qy	1054	YRDILFSG	SNDT--	SVSKDTP	SSSNDL	KNIYLLAS	STEOEREMK	----	1108					
Dh	1020	ykdllygs	ndtsdtl	gkqps	sndnlk	viyaesg	stegkekm	qkqakikl	inga	1079				
Qy	1109	KCSTERSAP	NULYSR	-QTMWEN	NGKTYI	WAGVCA	LTSKOKIA	KCYEKK	-PQKTE	PNEMIM	1166			
Dh	1080	tsqpyvr	tknsv	tkpqgt	lwenlak	dlmawm	calylk	enda	gatsak	leqndk	lkalw	1139		
Qy	1167	DEAKRKPK	POUYOT	YTNVTK	LENGST	PSPTQ	OASSDN	---	PTTLTH	PFKRPTRY	RRWEE	1224		
Dh	1140	deahrn	ptlekyq	tnvkl	edess	-----	aksndtl	ppctlk	nvtepr	fltrwhe	1199			
Qy	1225	WGESFCRE	RKRRLK	QIKLVDC	KVENG	DVGRSG	DGEACDS	I	STHND	SYVSPFC	PCGGRHC	1284		
Dh	1191	wgnsfere	akrlaq	ikhemedge	-kysgsg	geycce	elfakynv	ldls	-sasc	carp	1248			
Qy	1285	SSYKATLE	RKKIEH	KOSNA	YGQKTD	-----	ATRNNGT	FDKEKCT	LETLP	DDA	1333			
Dh	1249	rlvtytle	kkkete	qkayeq	qsknyen	eqkd	kcqts	nnma--	nfe	rtslg	asptaa	1306		
Qy	1336	KFLERL	NGCCKTK	KEY--	GGDD--	-IDPEKOS	KFFOTE	YCGC	OPKRTK	CONG	NGCVGL	1392		
Dh	1307	eflkl	--g	cknngyem	geahk	idfkn	ppckl	fkaahs	cdpr	ltyvk	qngnhc	-yssa	1363	
Qy	1393	NG-NC	DDSKS	IDA	KEIA	KMRS	STTDDV	MARY	SDMT	WTTFE--	GGDLK	DAQAHN	IKGRK	1449
Dh	1364	ngckch	nmk--	lteed	lknk	dprgn	liem	vsds	tnltf	hgd-----	cksg	lfgkrl	1417	
Qy	1450	DVMKCG	YVCYD	IC--	EQTIN	ERTOG	KEYIQ	IRAL	FKRW	ENFLE	DYKIND	KSHC	1507	
Dh	1418	devmcan	gvad	ctle	tklkn	qged	qkytkm	lkel	lwlw	efledy	ortfr	klkclck	1477	

[illegible]



```

OY 864 -----DETAKE-----TTEGS----- 875
Db 1452 lpsqflrmmytfqdyrdiclnltidiskqndvakakdkiqkflfskdgskspglstrqew 1511
OY 876 -----ATDP----- 879
Db 1512 ktngpelwkgmcaltkyvtddnkrkikndysdkvngsqngnpsleeaakpqflrm 1571
OY 880 ----- 879
Db 1572 lewgeefcaerqkkenilkdaeneinstqgcndakhrconcraygeyenkkkefsqtc 1631
OY 880 ----- 879
Db 1632 moflkanvpqdpkykyeykdqvplqgneylqkcdnkcscmdgnlsvspkexpf 1691
OY 880 -----TSLDVCPIYGVKLTIKDNEISLODA 903
Db 1692 gkyahkypeekcdcygqkhvpslpppppyvqgppeaprtvrvdcslv-klfkfdtmfsda 1750
OY 904 CSLKTYGNNLSRLGMRCV---TPSGEPT---SSDKNGAICVPPRRRLYIKKIVDMATKT 957
Db 1751 cglky-gktapsswkcipsdtksgagatqgksdgsiclprrrrllyvgklgwatal 1809
OY 958 ESPQASGSEASTSGSTTPPDSKEALLKAFVSALETFFLHNRKREKKAQAQGAGHG 1017
Db 1810 --pqegaaapshra-----ddlnafiqsaaletfllwydrkeekbpqg-qgsqga 1858
OY 1018 LPRVEE--GSPYDEDEK-LKEKIPDGLROMFTYLGDRDILKSGSDMTSVSADTPS 1074
Db 1859 lsglscstysddecdppdkllqgkllpdpflrmfyllqdyrrllvhggn--tsdsgntg 1916
OY 1075 SSNDNLKNIIVILLASGSTQERKMKRYKEI--KNEFKCSTERSAPNLVSHPOTWENNNK 1132
Db 1917 snnn--nlvleasgkmedmqklegqlpknsgtrlvprss--aqtpkwnnehae 1970
OY 1133 YTHHNMCALT-----SKDKIAGVKPKOKIENPENLME-----ANKK-----FKP 1175
Db 1971 slwkgymicaltyeknptlsargde--nklekdevyekffgsstadcgtastcltqlyk 2027
OY 1176 POYOTYNYKLDENSGSTSPPTOTQASDNTPTTLTHFVFRPTYPFMFEWGESFCERKK 1235
Db 2028 tqdyekykletdsg-----aktpsasdtpr-llsdfvtrpyfyleewgnfckktrh 2081
OY 1236 RLKQIKVDCXKVENGDVG-----RCSGDGACDSISTHDYSTVPFNCPCGKRCSY 1287
Db 2082 klagklheckveengsgstrgqilrtqysgdgaecnemlpkndgtvprdlekepcakrcsey 2141
OY 1288 RKMIRKRIEFPKQSNATGOOKTDATRNNGNTFDEKFECTLTETWDAKFLERLKNPCK 1347
Db 2142 tkwleskglekqkqayeqdk-dcvngsnkhndngfcteltsskakfkltl--gpek 2198
OY 1348 TTKREYGGDDIDPEKDSKTFPOHTEYCGPCPKFTNCONGNGCGVSLNGCDGKSIDAKET 1407
Db 2199 punvwctctfd---ddkftkhkdcopclksfnckkdecd-nsgytdcrnknsidacdi 2254
OY 1408 AKMRSSSTDVYVWVSNDNTNFEQDLEKDACQHANIFKGIKRDVWKCQVVCVDIQQTN 1467
Db 2255 engvdstv-lemrvsadsksfgngdlenacrsgagifegirtdemwckrnvgvyvckpkn 2313
OY 1468 INERTDQKEYIOIRALFKRWVNPFLBEDYKINDKISHCKKKEGSGKICNGCEKNSCLEK 1527
Db 2314 vngaeqknhlqirailvkrweyffedynklkhkshtrkngelspcd---kn--cvek 2367
OY 1528 WIEKTAIEMENIKRRFNDOYEKRDQDVWKSILEELJPKIYVNDQDVAIKLVEPENS 1587
Db 2368 wtdqkrkewkelterfkdgkxndsdadnvsfietllpqiadanaknkvylslsfngsc 2427
OY 1588 GCTLISNTQ--NNKENDAIQMLKRLGVAKNCPGK--PSEGEKOSDCKEPPPLPDE--- 1639
Db 2428 ggsaaneqngkngeykdaldcmllkklkldigceckhhqtsctesdprqpptleedetid 2487

```

```

OY 1640 ---EDQNPENITLBPPEKFCPP--TOPPEKGGETCGNKNEKKDEKE:SEEPKAEESGP 1694
Db 2488 ddleceeaaknmnm--pkicenvlktagedeg--c-----vpaeeiseepatdsgk 2535
OY 1695 AAEPAPPAESESETETNPEPPGTGPAPSRPAP--PTPPTPPLR:QADEPDSITL 1751
Db 2536 etpeqtpvlkpeeavpeppp-----ppgekapaipqpqpqpptqjldnphvltal 2589
OY 1752 QP-TIPGVALMLGSLAFLEFLKRTKASVGNLFQIOLIFKSDYDPTLKSSNRYIPYSD 1810
Db 2590 vstlswasvgjgfatfylvlkkktksvgnlfqllqpkpsdydipklspryipytsq 2649
OY 1811 RYKGTYYTMEGSDPEDKXAFMSDPTDVTSSEXEELINDIYVGSFKYTLLEVLE 1870
Db 2650 kyrgkrylylegdsgtds-gycdhysdltseseseeyemdnolvygspkyxcllevle 2708
OY 1871 P-----SGNNTTASGKNTPSDTRNDIIONDGIIPSKITDYIEMNOLKKEFIISN 1916
Db 2709 psgmnttasgnttasgnttasgntktpdctqndiqndgipsskiltidewnglkdeifsq 2768
OY 1917 MLONDPNDVNDYTSNGSSTNTNTTTSKRNVNDNTNTTMSKDNMEIULLPSIHGONLY 1976
Db 2769 yiqsepnlepn-----mlgyuvdnthpctshnvee:pfimsindrnlf 2813
OY 1977 SGEESYNY-----NMVN-----SMNDIPINDNNVYSGIILINDSLSGKPI 2019
Db 2814 sgeeynydmfnsgnnpnlisdtsnmdsltsnbnhpndkndlysgjllindalsqn-h-1 2872
OY 2020 DIYDEVLKREKENELFGE--NTRKTSQNVAKTNSDPINQOLELFEKALDRRDMCEKK 2078
Db 2873 dlydenmlkkeknelfgtkhthtntnyakpardpietnglnlflkxldirndmcekk 2932
OY 2079 NKEDIILKKEPMKNENINNSGKTYNSDNKPSHNHVLNTDVSIOQDWNPKTKNEITND 2138
Db 2933 nherljpklkelw--enehsgdi--nsglpsgnhvlntdvsiqldndrpklnelntmd 2988
OY 2139 TNOCKSTMOTIIDLEKYNDPYYPFEEDDIYHDVDVKKSSMDIYVDBHNVTSNNMV 2198
Db 2989 tnpdskntmdlliddlekynepyydieddillyhdvdeksmddilyvhnvntnmndv 3048
OY 2199 PYKMHENIIVN 2210
Db 3049 pckmhlemivn 3060

RESULT 6
Y77905
ID Y77905 standard; Protein; 3060 AA.
XX
AC Y77905;
XX
DE 13-JUN-2000 (first entry)
XX
DE Plasmodium var-7 polypeptide.
XX
KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW DAB; Stialic Acid Binding Protein; SAbP; malaria; vaccine; immunisation;
KW protozoacide; var-7.
XX
OS Plasmodium sp.
XX
PN US5993827-A.
XX
PD 30-NOV-1999.
XX
PF 07-JUN-1995; 95US-0487826.
XX
PR 10-SEP-1993; 93US-0119677.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Sim KL, Chltnis C, Peterson DS, Su X, Welliams 'E, Miller LH;
XX

```

DR WPI: 2000-194198/17.  
 DR N-PSDB: 298287.  
 XX  
 PT Isolated protein binding domains from Plasmodium vivax and Plasmodium  
 PT falciparum erythrocyte binding proteins useful for vaccinating against  
 PT malaria -  
 XX  
 PS Disclosure: Columns 109-124; 93pp; English.  
 XX  
 CC The invention relates to ebl-1 polypeptides that are encoded by the DBL  
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially  
 CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid  
 CC Binding Protein (SABP), which are soluble proteins that appear in the  
 CC culture supernatant after erythrocytes infected with malaria release  
 CC merozoites. Immunohemical studies indicate that DABP and SABP are the  
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy  
 CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be  
 CC used to vaccinate against malaria, especially caused by P. falciparum.  
 CC Immunization with the polypeptide provides effective protection against  
 CC malaria. The present sequence represents the var-7 polypeptide.  
 CC  
 XX  
 SO Sequence 3060 AA:

Query Match 39.7%; Score 4802; DB 21; Length 3060;  
 Best Local Similarity 36.2%; Pred. No. 3.1e-281;  
 Matches 1156; Conservative 277; Mismatches 631; Indels 1128; Gaps 83;

QY 4 SGGSGGTODEAKNVDFEGQKVNDEVHNGEAKNVSELKSLASLIGETAFVTKSQOT 63  
 DB 12 aagddeddesaklmfrlgkdydkveakegkglqrls-----eakfknesdp 65  
 OY 64 ES-----KYTELIEANSKRNPCK-----KDGKGNVDNRSVYEQAGYKMKKC 107  
 DB 66 qtpedpcedlthkyhtntlt-n-vinpcadrsdvrfdsdyggqctmrllksqgdnkg--- 121  
 OY 108 SNGMTCAFRRLALCNKNFPMNSNDSSKAKHDLAEVCMAYKEGESIKTHRYKDYKY 167  
 DB 122 ----acapyrrllhvcddqleqleplklnt-nhllvdvcmaakefegslctgdykqygt 176  
 OY 168 PGSPFPCMTLARSFADIGDILRGDLYLGKKNKKQNGKEFEREKLQOKLEIKTKITDN 227  
 DB 177 gdspsqgclctmlarsfadlgdlvrgdrlylgnpgeik-----qrqqlennlltfikgklyek 231  
 OY 228 LKQDEAKRYNGDEDPNPFYKREDMWTANRETVGAMTCSKELDNSSFRATCNDTQGP 287  
 DB 232 lngaea--ryg--ndpefiklredwcanreclwkaalcna--gnlyflactn---rg- 281  
 OY 288 SQTNNKCRCDKCGAMAGKPKAGDGVITVPTFVDPVYQVYLRFEEWAEDEFCRRKKKLE 347  
 DB 282 ertkygcrnddq-----vptyfdyvpqylrfeewedecrkhkklk 325  
 OY 348 NLEKQCGKRSDEYRYSRNGYDCBOTISRKGVKMGKCTDCEFPACGSENMIDNOK 407  
 DB 326 dvknrcygdckdedyrcsngygcetkralgklyrgkqglsclycnpydvdlmqke 385  
 OY 408 QPDRQ-KKTYKEI-----SDGGGRKKRAVGG--TTKYEGEKSFEYELKNDGTVDAF 458  
 DB 386 qfdqgkkydeekkyengagaagstrqkrdagyltctmgydekkfydelinkseyrtvcdkt 445  
 OY 459 LGLLNNEKACKDITD--GKINFEVNSGGVGGSGGTSGASTNDENKGFYRSEYC 516  
 DB 446 leklsneelctkvdeeggtldfkvn-----sdsstgasggtlveagstlyreyk 496  
 OY 517 QPCDDCVQ--HKGC--NOMERKTKVKKMRWSKLYK---INGKMLLLKSLAKVKKDMIT 569  
 DB 497 qpcygcgvkynnggsaneweeekn-gkcskgllyepkpkdegtltilksgkphdd--- 552  
 OY 570 LKKMKKEECLTONSDSGSVGVTVTGASGNSKEKLEIDEKCKYHNHYOKVNNQGEVEE 629  
 DB 553 leeklnkfcdckngdclnsqsgstcgsqgnsqdelyeekcykgegvvynghdedee 612  
 OY 630 DDELKAGAGGLCTLPNPKKNKEVSEAKSONNHADIOKTFHFYUVAHMLKDSIHMRTK 689

DB 613 dyenwnaagylcilkngkknkeegntsekepedeqtktnpfyfywvahnmlkdslnhk-k 671  
 OY 690 RLKCSISGKRTMKC-RNCCNKKCCDCEKFWYKQKETEMKPRIDHDKRTQSGIEGYFTTLE 748  
 DB 672 klqrcilqgnrtikcgnknkcndcecfkrwlrtqkhdwqklyqhktlqnlyrgsgsdhae 731  
 OY 749 LI-----LKTQFKED-----TEENTENSLDAEAEELKHLQKILKLENNLAV 793  
 DB 732 lipdhdvlyqnlqgeellkgsdsedaseeksenldeaeelkhlreliscednngas 791  
 OY 794 VNAG-TEQKTLMDKLNHLELNDATKCCDCLPEEDK----- 828  
 DB 792 vggvteqknmdkllyekdeadclleheedeekkgdgnecieegenfrynpsge 851  
 OY 829 ----- 828  
 DB 852 sgnkrypylankvaygmhkhaktqlaasragrsalrgdsiaqfkngrngstlkqjckln 911  
 OY 829 -----SKRGADPS-----PDIFIP----- 843  
 DB 912 enysndrsgnsgrpctygdghgvyrmrgtewsnieggkqtskynvllprrehmctsn 971  
 OY 844 ----- 843  
 DB 972 lenldvgsvtknkashelldvgqlaaktadaaelkrykdqgnlqltdpjqgkqeamcr 1031  
 OY 844 ----- 843  
 DB 1032 avrfsfadlgdlirgrdmwdeksdtdmetrlitvfnlkexhdgikdnpytgdskkp 1091  
 OY 844 ----- 843  
 DB 1092 aykklradwneanhqvwramkcatkglicpmpvddyipqrlwmcmewaewykagsqe 1151  
 OY 844 ----- 843  
 DB 1152 ydkllkicadcmxgdygctgdvdcgkcaacdkyeeleakwneqwrklsdkynllylq 1211  
 OY 844 ----- 843  
 DB 1212 aktstnprtlvgddpdygmvdflrlphkasiaarlvakraagptelaaaapltry 1271  
 OY 844 ----- 843  
 DB 1272 staagylhgeiygycgeqtcfeekkhgatsstlctkenkeytlfqpppeyatadclnrs 1331  
 OY 844 --RPEEKED----- 850  
 DB 1332 qteepkkeeenvesaaklvekilegkngrttvgcnpkhesypdcknldishdgacmp 1391  
 OY 851 -----DEN-----EDDEDEVRD--- 863  
 DB 1392 prrkcllylahesqenlktddnlkdaflktaaeelflsqwyakakndseaklldrgl 1451  
 OY 864 -----DETRAKE-----TTGES- 875  
 DB 1452 lpsqflsmymtyftgdyrdicntdlskqndvakakdklgykffiskdskspsglsrgew 1511  
 OY 876 -----ANDT----- 879  
 DB 1512 ktngpelwkgmlcaltkyvtldtnkrklkndysdkvngsqgnpsleefaakpqflrmw 1571  
 OY 880 ----- 879  
 DB 1572 lewgeefoaerqkenlkkadacnelnstqgcndakhrcngacragyevankkkefsqgt 1631  
 OY 880 ----- 879  
 DB 1632 nntvlkanvpqdpeykygyekdgvqplqgneyllyqkcdmknksacmgnvlsvapkckpf 1691  
 OY 880 -----TTSLDVCPIVGKVLTKDNESLQDA 903



D	b	1692	gkYahkYpekcdcygqkvhvpslpppppyqbpqpeapltvdcslv-ktlctdnfda	1750
Q	y	904	CSLKYGGNNRLQRCV---TSPGEPT---SSDKNGAICVPPRRRLYIKKIVDMATKT	957
D	b	1751	cglky-gktapsawcipcstckagagatctgksdgsdsciclprrrrllyvgklqewatal	1809
Q	y	958	ESPASSSEKSSSTSTTPPDCKEALIKAFVESALETFFLIMHRYKEEKAANOAGCHG	1017
D	b	1810	--pqgeaapshera-----ddlrnafigsaaletfflwdrykeekbpqg-dgsqqa	1858
Q	y	1018	LPRREE--GSPKXDPEDK-LKEKIPDGLROMFTLDGYRDLFSSGNDTTSVKOPIS	1074
D	b	1859	lsqcltysddeeppoklllqngkippdcltmytllqydrilvlyhgn--tsdsagnv	1916
Q	y	1075	SSNDLKNVILLAGSTGEOREKMKRYKEI--KNRKCSTERSAVNLVSHQTMWENNGK	1132
D	b	1917	snm---nvlssngkemdmgkqekleqllpknggtrlpvpxs---aqtpdkwnehae	1970
Q	y	1133	YIHGMVCALT---SKDKIAKVERKPOKLENPENLME-----ANKK-----PKP	1175
D	b	1971	slvymicaltyeknpdtsargd---nkiekdeveyekffgsrtdkhygtsaptryk	2027
Q	y	1176	PQYQYTWKIDENSGISPRITQOASDNTPTTLHFKRPRYFWMFEKMEBSCREKKK	1235
D	b	2028	tyqdyekvklctedsq-----aktpssadclp-lldvltprtyflieyewgnctkikh	2081
Q	y	1236	RLKQIKVDCAYENGMDVG-----RCSDDGEACOSISHDSTVSPFNCGCGCHKSSY	1287
D	b	2082	klagikheckveenggsrrrtgyltrgysdgaecnemipkngdvtpldekscapecasy	2141
Q	y	1288	RKWIERRKIEFHQSNANAYGOOKTDATRNNGNTFDEKFTCLTEWDAKFLERLKNBPCK	1347
D	b	2142	rkvlsskygkfeqekayeqqg-dkcvngsnhdngfcelitstaskafliktl--gpck	2198
Q	y	1348	TNKRGGDDIDFEKDSKTRQHTHEYGCPCKFRFTNQONGCVCYSLNGCDDGKSIDAKEI	1407
D	b	2199	pnnvegtrftfd---ddctflhtkdcpcjclfsvnckkced-nskytclernknsldatd	2254
Q	y	1408	AKMRSSTTDVHNVSDNNTFTFEGDOLKQACOHANFKIRKDWKCGYCVGVIDCEOTN	1467
D	b	2255	engdvstev-lcmvssadsksgfrngdlenacrgagflegirtdemkcrcnvogyvccle	2313
Q	y	1468	INERTDKEYIQTALFRKMWENLEDYNNKINDKISHCIIKKGSKCJINGCEKNSCKLEK	1527
D	b	2314	vngsakghhliqtraltvkrveyfedyakikhhkshirkingelspcl---kn--cvck	2367
Q	y	1528	WIEKKIEMENIKRRFNDQYENKDDYVWVKSLEBELPKIAVNVDDQNVYIKLCVEFKS	1587
D	b	2368	vwdckrkrkewkelerfdqgkhdnsdddvzsfletllpqldanakhvklakfgnsc	2427
Q	y	1588	GCTLISTNTO--NNKENDADIDCMKTKGVAKNCPG---PSCGKSDCKRPPPLPDE--	1639
D	b	2428	gcsasaneqkngcykdalcmklkldkligceekhhqtsdctescdpqpqriledelid	2487
Q	y	1640	---EDQNEENTLEPPKFCPP--TOPPEKGGETCGNKEEKDEKKEESEPAAKEESGP	1694
D	b	2488	ddleteeaknmh--pkicenvlktagdegsg-c-----vpaenseapatsgk	2535
Q	y	1695	AAEPAPATASEETENFPEPPECTGAAPSPAP---PTPTPPPLRLQADDEPDPSTIL	1751
D	b	2536	etpqrtpvlrpeeeavpeppp-----ppqekapaplpqpqrtpcpqlidnphvltal	2589
Q	y	1752	QT-TIPPGVALAGSLAPLRLKKTAKASVGNFOILOIPKSYDIPTLKSSNRYIPIYSD	1810
D	b	2590	vtclslawsvigfatfyfyrlikkktlssvgnlfgllqprksodydipklsprtylpycsq	2649
Q	y	1811	RYKKGTYIWEGSDDEBKAFYMSDMDVTVSSESEYEELDINDIYVGSPPKTKTLEVLE	1870
D	b	2650	kyrgkrylylegsgyds-gytdhysdltsseseyeamdindiyppspkyltllievle	2708
Q	y	1871	P-----SGNNTTASGKTPSDTRNDIONDCIPSKTITDDEMNQKKERFSN	1916
D	b	2709	psgnntlaasgnntlaasgnntlaasgktpsdtdndipndipsskltidewnglkdeifsg	2768

QY	1917	MLQNPDPNDVPDIDYSGGSSINTNTITTSRRINVDNNNTNTMSRDNMEN.LLPSIHGDNLY	1976
Db	2769	ylgsepntlepn-----mlgynvdnaphptishhveek2flmslndrnlf	2813
QY	1977	SGEEYSYNNV-----NMVN-----SANDIPINDNNVYSGIDILINDSLSGKPI	2019
Db	2814	sgeeynydmfmsgnnpnlisdstcsmdslltsnmhsygnkndlysgidlindalsgnh-i	2872
QY	2020	DIYDEVLRKRENELEFGE-NTKRTSPQONAKTYSNPDIHNOLELFHKHLDRRHDCERWK	2078
Db	2873	diyemlkrkenelfgkthkhtkhtnlynvakparddplnqnlfnhka.ldrhrdmcekwk	2932
QY	2079	NKEDIKLKREEMKKEINNNSGRTYNSDNKPSNNHVLNIDVYSIQIDMLNPKTKNETIWD	2138
Db	2933	nnherlplkkelw--enethsgdl--nsglpsgnhvlntdvsyqldmtnpkLkneiitmd	2988
QY	2139	TNOOKSPMDITLIDLEKYNDPPYYDFEYDIIYHDVDVEKSSMDIYV'DHNVTSSNMDV	2198
Db	2969	tnpdkstndctldldlekynepyydfyeddilyhdvdveksmdily'dhnnvtnnmndv	3048
QY	2199	PTKMHIEKNIVN 2210	
Db	3049	ptkmlhlemiaivn 3060	
RESULT	7		
W00384			
ID	W00384	standard; Protein; 2913 AA.	
XX	W00384;		
AC			
XX			
DT	21-FEB-1997	(first entry)	
DE		Plasmodium falciparum erythrocyte membrane protein.	
XX			
KW		Plasmodium falciparum; erythrocyte membrane protein; malaria; detection; identification; treatment; prevention; parasite.	
OS		Plasmodium falciparum MC type.	
XX			
XX	Key	Location/Qualifiers	
FM	Domain	62..394	
FT		/label= Duffy binding ligand domain 1	
FT	Region	607..648	
FT		/note="Cysteine rich motif"	
FT	Domain	839..1272	
FT		/label= Duffy binding ligand domain 2	
FT	Region	1482..1527	
FT		/note="Cysteine rich motif"	
FT	Domain	1706..2005	
FT		/label= Duffy binding ligand 3	
FT	Domain	2102..2349	
FT		/label= Duffy binding ligand 4	
FT	Region	2354..2398	
FT		/note="Cysteine rich motif"	
FT	Domain	2450..2475	
FT		/note="Putative transmembrane domain"	
FM			
PN	W09633736-A1.		
XX			
PD	31-OCT-1996.		
XX			
PF	26-APR-1996;	96WO-US05798.	
XX			
PR	27-APR-1995;	95US-0430908.	
XX			
PA	(AFFY-) AFFYMAX TECHNOLOGIES NV.		
XX			
PI	Baruch DI, Howard RJ, Pastoske BL.		
XX			
DR	WPI; 1996-497376/49.		
DR	N-PSDB; T41852.		





Db 7 aagdddlledesakmfdatlgkdydkvkeakegkglqgrls-----eakfeknesdp 60  
QY 64 ES-----KYTELEANSKRNPCK-----KDGKGNVDPRFSYKQAGYDNKKMKC 107  
Db 61 qtpedpoldhkyhtnvtln-vlnpcadrsdvrtfsdeyggcctnrlkdsqgdhkyg--- 116  
QY 108 SNGMTCAFRLHLICNNKFPNMNSNDSSKAKHDLAEVMAKYEGESIKTHYPRKYDSKY 167  
Db 117 -----acapyrrllhvcdqqlqepikltent-hnllvdvcmaakfeggsitcdyprkyqcty 171  
QY 168 PGSDPMTMLABSPADIGDIIRGRDLYLGKMKKQKQNKQKKEFERELBECKLEIRKIHND 227  
Db 172 gdspsqctcmrlarstfadlvrgdlylgnpqelk-----grqqlennlltlltlyklyek 226  
QY 228 LKDEAOKRYNGDDEPNFYKLEEDMWTANRETVNGAMTCSKELDSSYFRATCNDTGGCP 287  
Db 227 lngaea---ryg--ndpefklrtedwtenretvkalcnaw--gnlyfhtcn---tg- 276  
QY 288 SQTINHCRCDDKGNAMAKPRAGDGVITVPTFYDYVPOYLRFWEEMADECRKKKKKLE 347  
Db 277 ertkyccrncddq-----vptyfdyvpqylrtfweeadeferckknkkik 320  
QY 348 NLEKQCRKQKSDERYCSRNGYDCEOTISRKKYVRMKGCTDCPFACGSENNMDNRK 407  
Db 321 dvkncrcykdkeddyccsrngydccklraigrlykqclsclyacnpyvdlmgkø 380  
QY 408 QFDRQ-KRYTKEI-----SDGGRKKRAVGG--TTKEYEGEKSYREKLKNDGYGVDAF 458  
Db 381 qfdqykkydeeklkyengasgsgrqktrdaagcttltngdygekkyfdehlnkseyrtdkf 440  
QY 459 LGLINNEKAKCDITD--GKATNFKEVNSGGVGGSGTSGATDENKGTFFYRSEYK 516  
Db 441 leklsneelctkvkdeegtldfknv-----sdstsgasglvnesggtfyskyc 491  
QY 517 QPCDGCQV--HKGG--NOMERKTQVKKMKRMSKLYKPR--INGKVVLLKSLKLVVKDMIT 569  
Db 492 qpcpyccgkxvlnpgssnweekn-gkcxsgklyerprdkqegttlltllkqkghd--- 547  
QY 570 LKRWKRECLTQNSSDSGSVVTTGASGNSSEKKELDENKCYKHNENVOKVNGOGEVEE 629  
Db 548 leeklnfcdckngdltngsgsgtsgsgsgngsfqelyeekcykgevlvngvndedee 607  
QY 630 DDELKGAAGLICILPNPKKKEVSEAKSONNHADIOKTEHDFEYVWVANMLKDSIHMTK 689  
Db 608 dyenlvknaagstcllknqkknkeegntsekerdeqltnfffyuwalhmkdslnhk-k 666  
QY 690 RLKSCISGKTMKC--RNGCNKKCDCEKVMVKOKETEMKPRIDNHKTCQGTGSEGYTTLE 748  
Db 667 klqfclngnrlkcgknkcnndcecfkrwlqkxkdwqklyqhhtkltgnlkagrgsdntae 726  
QY 749 LI-----LKQFLKED-----TEENTENSLDAEAEELKHLQKILKLENNENLAV 793  
Db 727 llprdhdyvlygnqeeellkgsedaseeksenldeaeaelhlhreltisednngas 786  
QY 794 VNAG--TEOKTILMDKULNHELNDATKCKDCPLREEDK----- 828  
Db 787 vggvteqklnmkllynxekdeadclleheedeekekygnecieegenfryncpsø 846  
QY 829 ----- 828  
Db 847 sgnkrypvlankvaygmhkhaklqjastragsalrgdlsiaqfkngrngstlkgickln 906  
QY 829 -----SGRSADPS-----PDIFRP----- 843  
Db 907 enynderagnsgspctgkdqdhgvyrmrlgtewenlsegkkytsknyvllprtrehmctsn 966  
QY 844 ----- 843  
Db 967 lenldvgsvtknkashnsllygdvqlaaktadaaelikrykdqgnlqlldprlgkqdeamcr 1026  
QY 844 ----- 843  
Db 1027 avrystadlqdlgrdgmwdeksstcmetrlltvcfnlkekhdglykdnprkytgdeskkr 1086

QY 844 ----- 843  
Db 1087 aykklrtadwneanrhqvrnankcatkglicpmpvddyipqrlrwmteaweyckagsqe 1146  
QY 844 ----- 843  
Db 1147 ydklklcadomsksgdykctgtgdvdcgckaackkykeeleekwneqwrkltsdkynlllylq 1206  
QY 844 ----- 843  
Db 1207 aktstcpgtrvlqddpdyqmwdfllprlnkasiaarvlvakraagsptelaaaapltry 1266  
QY 844 ----- 843  
Db 1267 staagylnqelgygqceqctgfceekkhgatssttkenkeytfkqppeyatadclnrs 1336  
QY 844 ---RPEEKED----- 850  
Db 1327 qteepkckkeenvesacklvekllqgkngrtlvtgecnpkesydwcdknnldishdgamp 1386  
QY 851 -----DEN-----EDDEDEVRO--- 863  
Db 1387 prqklcllylahesqtenlktldnldafiktaaaetflswgykkskndseakltdrgl 1446  
QY 864 -----DETAKE-----TTEGS----- 875  
Db 1447 lpsqflzsmmytfqdyrdicintdstskqnvdavakakdklqkfiskdgskapsqslrqew 1506  
QY 876 -----ADTF----- 879  
Db 1507 ktngpelwkgmllcaltkyvtldtnkrlkndysdkvngsqngnpsleefaapqlrwm 1566  
QY 880 ----- 879  
Db 1567 lewgeefoaeqrkkenllkdaacnelnstqgcndakhrcnagcrayqeyvenkkhefsqct 1626  
QY 880 ----- 879  
Db 1627 nntfikanvqpdreykyeykdgvdprlqngneyllqkcdmkkscsmggnvlsvapkckrf 1686  
QY 880 -----TSLDVCPIVGKVLTKDNESLQDA 903  
Db 1687 gkyahkyrekcdcygkxhvpstlppppvpqpqpearlvtvdcslv-klfkcdmfnfsda 1745  
QY 904 CSLVYCGNNSRLGRCV---TPSGEPPT---SSDKNGACVPPRRRLYIKIYDMATKT 957  
Db 1746 cglky-gktapswwkclpsdtksgagatctksgdsqslcprtrrrllvygklgewatal 1804  
QY 958 ESPQSGSEASTSGSTTPPDSEALKAFAVESAAIEFTFLMHRKBEKKAVAOEGAGHG 1017  
Db 1805 --pqgegaaphrsa-----ddlrfnafisaatctflwdtykeekkpqg-dgsqga 1853  
QY 1018 LPRVBE--GSPETPDEK-LKEGKIPOGFLKQMTYTLGDIYDLFISSSNDTTSVKOTPS 1074  
Db 1854 lsqtlstysddeepbpkllyngkrlipdflrmlytlylgydydllyvhgn--tsdsgntng 1911  
QY 1075 SSNDNLKNIVLASSTGEOREKKNKYKEI--KNPRCSTERSAPNLVSHQGTWENNGK 1132  
Db 1912 snm--nlvleasgnndmqklqekleqllpkngsrplvpks--aqprdkwneae 1965  
QY 1133 YIWHGWCALF-----SKDKIAKGYEKKPKOKIENPENLME-----ANK-----PKP 1175  
Db 1966 slwkgmicalalyteknptdsargde---nkkekdeyvekffgstackhgyastprltyk 2022  
QY 1176 POYQYTNVKKLDENSGTSBRTTQIOASSDNTPTTLTHFVVKPRTYRFRMEGESCRCERKK 1235  
Db 2023 tgydyekvklledtsq-----aktpsaasdlp-llsdfvlrppryrleewgnqclckrkxh 2076  
QY 1236 RLKQIKYDCKVENGDVG-----RCSGGEACDSDSISTDHSHVSPFNCPGCGKHGSSY 1287  
Db 2077 klqaklneckveengsgstrrglctryqsgdgeacnemlpkngdflvpdllekpscakpssy 2136



OY	630	DDBDELKAGAGGLCITPNPKNKKEVSEAKSONNHADIQTEFHDEFFYVWVAMLKDSIHWPTK	689
Db	615	dyenvknagglc1llknqkknkeegntsekepeidqktfnrfflyvwalmlkdsihwk-k	673
OY	600	RLKSCISDGKTMKC-RNGCNKKCDCEFKWVKQKETEWPRIKHDFKTOEGIPBGYFTTLE	748
Db	674	klqrcqlqgnrtikcgnnkcmmndcecfkrwtlqtqkdwgkltvqfktqnlkxrgsgsdntae	733
OY	749	LI-----LKLOFLKED-----TEENTEMSUAEEAEELKHLQTLKENNNNAV	793
Db	734	l1pfthdyvlqynlqgeeflkqdsedaaseeksensldaeaeelkhlreliesednngas	793
OY	794	VNAG-TEQKRTLMDKLNLHNEULDATKCDKDCUPEDEK-----	828
Db	794	vvggyvtqeknlmdkllnyekdeadielhelndeeeekexydgndecleegenfrfynpcsg	853
OY	839	-----	828
Db	854	sgnkrypvlankvaygmhbhkkclqlaesagrsalrgdislaqfkngrngstllkgkglcn	913
OY	829	-----SNGSRADS-----	843
Db	914	enysndsrzgnsgspcylgkdqdhgvrzmr1gltewsnlegkqtcsyknvflfprrehmctsn	973
OY	844	-----	843
Db	974	lenldvsgvtckndkashsllgvdqvlaactaaelkrryadqmg1qltclp1qgkqdeamcr	10333
OY	844	-----	843
Db	1034	avrysfaelqgl1lrgdmwdeksstametrll1tvtfnikekhdg1kdnpytgdesskpr	10939
OY	844	-----	843
Db	1094	aykkl1radwganrhqyvrwamkcaclkyllc1cpmpvd1y1pqr1rwmteaweycksgsge	11533
OY	844	-----	843
Db	1154	ydkllk1loadcmsskgdgc1kgdvd1ogkckaaedckyeek1ekwneqwrk1skynll1y1q	12133
OY	844	-----	843
Db	1214	aktetnprtlvlgdddpdyqgmvdfltp1hksaslaarvlvkrraagsp1e1aaapl1c1y	12737
OY	844	-----	843
Db	1274	staagy1lqel1y1ggcqe1q1fcekknhgastst1ckenkey1fkqpppeyatacd1nrs	13333
OY	844	--RPEEKED-----	850
Db	1334	qteepk1k1keenveesacl1vek1legknr1trtvcgcm1pksyp1dwcdk1mid1shbgacmp	13939
OY	851	-----DEN-----	863
Db	1394	pr1rql1cl1y1ahesq1ten1k1td1n1k1daf1k1taaaet1lswqy1k1skndseak1ld1r1g1	14533
OY	864	-----DEETAKE-----TEEGS-----	875
Db	1454	lpsqfl1rsmmy1t1fgd1r1d1c1n1d1sk1gn1d1vak1ak1d1q1k1f1fs1d1g1k1sps1q1srg1w1	15133
OY	876	-----ANDT-----	879
Db	1514	ktngpel1k1gm1lcal1tky1vcd1tn1k1r1k1nd1y1dk1vng1sg1nps1leef1aak1p1q1l1rwm	15737
OY	880	-----	879
Db	1574	1ewgeefcaerqk1ken1lk1dace1n1st1g1cnd1ah1nrc1ag1cray1gey1enk1k1ef1sg1t	16333
OY	880	-----	879
Db	1634	nmf1k1a1nvp1q1p1dey1k1gyey1k1d1qv1p1l1q1n1ey1l1q1k1cd1nk1c1sc1md1gn1v1s1vp1k1e1k1pf	16933

QY	880	-----TSLDVCPIYGVKLTIDNLSLODA	903
Dd	1694	gkyahkyepkcdcygqkhvpslpppppyvpqpgeapbtvdcslv-kllfdlnlfisa	1752
OY	904	CSLAKGNGNSRLGRCV---TPSGEPT---SSDKAGALCVPPRRRLYIKKIVMAATKT	957
Dd	1753	cglky-yktapsavkcclpsdtksagatctgksdsagsclpprrrrlyvgklgwataal	1811
OY	958	EPPQASGEAASSTGSTTPPDPSKEALIKAFVESAALETFFLMHRYKEEKKAVAOAGHG	1017
Dd	1812	--pgsegaphrsra-----ddlrnatfgaaletffldwrykeekkpqg-dgsqga	1866
OY	1018	LPRVEE--GSPEYDEPK-LKEGIKDPLGRKFQMFYLGDYRDLFSGSDMTSVSKOTPS	1074
Dd	1861	lsqtlstysddeelpokllqngkfipdfilrmfylelgdyrdllvhgn--tsdsngnt	1918
OY	1075	SSNDLKNTVLTLASSPNEOEEMKNKKEL--KNFKCSTESAPNLVSHPTWMENNKG	1132
Dd	1919	snnn--nlvsasgnkedmqkfqekleqlprpngtclpvpxss--aqtpdkwnenae	1972
OY	1133	YIWHGMNCAL----SKDKIAKAEVKPKQENPENIMADE----ANK------PKP	1175
Dd	1973	sivgmicalyteknpdtsargde--nhkledeveyekffgstadkhgtastptglyx	2029
OY	1176	PQYOTNVKLDENSCTSPRTTOAOASDNTPTTLTHFPKRPRPYFRMVEEGESFCREKK	1235
Dd	2030	tgydyekakledts-----aktpsasdclp-lhsafvlrppryfyrlsewgqnfckkkh	2083
OY	1236	RUKOKVXCKVENGDVG-----RCSGGACDSISHTDYSPVSFNCPCCGKHGSY	1287
Dd	2084	klaqlkhecvkeenggsrrrgiltqysagdeacnmllphnqglvypdlekpcahpssy	2133
OY	1288	RKMIEKKIEFHKOSSNAVGOOKTDATRNNGTFEDKEFECTLETPWDAKFLERLKNPGK	1347
Dd	2144	rkwleskgtekekakeayegk-dkcnsgnkhdngfcetlitsakadflctl--gpkc	2200
OY	1348	TNKKEYGDDIDFEKDSFTFOHTEYCGBPCKFTNCONGCVSGLNGNCDGKSIDAKEI	1407
Dd	2201	pnnvegkllifd---dktlfnhtkdcpcjklfsvnckkdec-d-nskgtdcornksidatdi	2256
OY	1408	AKMSSTTDVMMRYSDDMDTFFGGDDLKMACQHANIFKGIIRKDWKCGVYGVDICEBTN	1467
Dd	2257	engyvdstv-Iemrvsadsksgfngdglenacsraglfefirfdewkccrnvcgvvvckpen	2315
OY	1468	INERTDGKEYIQIALRKRWNVNLFEDYNINDKISHCIKKEGSKJINGCEKSNKSCLEK	1537
Dd	2316	vngeakqghihqitalykrwvevfefedynklkikhshrtkingeispcl----kn--cvek	2369
OY	1538	WIERKIAEWENIKRFENDQYENKDDQDPNVKSIILEBLIPKAVVNDQDNVIKLCEVENS	1587
Dd	2370	wvdqrfkrewekteiterfkdygkndnsdddavrsleflllpqildcanakhkvilkfjgnc	2429
OY	1588	GCTLISNTQ--NNKENADICMLKTKGYAKNPGK---PSEBKOSDCKEPPLPDE---	1639
Dd	2430	gcसानेनगकयगाकािदमिल्लिकदिगेकेहभतसेदसदेरपुलेदलिद	2489
OY	1640	---EDONPEENTLEPPKFCPPPT--TOPBEKGETGNKEEKKDKKESEPFKAEESGP	1694
Dd	2490	ddieteekkhmm--pkiicenvlktgagedegs--c-----vpaenseepaatdsqk	2537
OY	1695	AAEBAPTAESEETNFPEPFGTGAAPPSPAP--PTPTPPLLRQAODEPDSITIL	1751
Dd	2538	etpqetplvkpeeaavepppp-----ppqeakarplrpqpprcpqcqllidnpvltal	2591
OY	1752	QT-TIPEGVALLAGSIAFLK 1772	
Dd	2592	vtslawsvgiyfalfcyfylk 2613	
RESULT	10		
ID	y77904	standard; Protein: 2710 AA.	
XX			

```

AC Y77904:
XX 13-JUN-2000 (first entry)
DE P. falciparum Proj3 binding domain polypeptide.
DE DBL gene: Duffy-binding like gene; ebl-1: Duffy Antigen Binding Protein;
KM DABP: Sialic Acid Binding Protein; SABP: malaria; vaccine; immunisation;
KW protozoacide; Proj3.
XX Plasmodium falciparum.
OS US5993827-A.
PN 30-NOV-1999.
XX 07-JUN-1995; 95US-0487826.
XX 10-SEP-1993; 93US-0119677.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA Slim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
PI WPI: 2000-194198/17.
DR N-PSDB: Z98286.
XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT falciparum erythrocyte binding proteins useful for vaccinating against
PS malaria -
XX Disclosure: Columns 79-92; 93pp; English.
XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
CC Binding Protein (SABP), which are soluble proteins that appear in the
CC culture supernatant after erythrocytes infected with malaria release
CC merozoites. Immunochemical studies indicate that DABP and SABP are the
CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
CC used to vaccinate against malaria, especially caused by P. falciparum.
CC Immunisation with the polypeptide provides effective protection against
CC malaria. The present sequence represents the Proj3 binding domain
CC polypeptide.
XX Sequence 2710 AA;
SO

```

Query Match 26.4%; Score 3197.5; DB 21; Length 2710;  
 Best Local Similarity 30.6%; Pred. No. 2.8e-184;  
 Matches 834; Conservative 244; Mismatches 569; Indels 1075; Gaps 74;

```

OY 4 SGGSGTODEDEKAVLDEFGQKVDHEVHGAEAKNVYSLKSGSLASILGETAFVYSMOT 63
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 14 aagggddiedesakhmfdrigkdvkveakeerkgylgrls-----eakfenesp 67
OY 64 ES-----KYTELLEANSKRNPCK-----KDGKNDVDRFSVKEQAGYDNKKMK 107
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 68 qfpedpcdlhkyhenvtn-vlnpcadrsdvrfdsdeyggqcthmrlkdsqgdhkg--- 123
OY 108 SNGMTCAPPRRHLCNKNPNMNSNDSSKAKHDLAEVCMAMAKYEGESKITTPRYDSKY 167
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 124 ---acapyrrlhvcdqnlqleipiklnt-hnllvdvcmakfegqsitgdpykyaqly 178
OY 168 PCDSPMCTMLARSPADIGDIRGDLVGNKKKONGKETREKLEOKLEIFPKIHDN 227
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 179 dspsgicmlarsfadiygrdlylgnpqel-----grqlenniklftgkiykp 233
OY 228 LKDKBAQRKRYNGDEDPNFYKLRDEWMTANRETVGAMTCSKELDSSYFRATCNDTGOCP 287
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 234 lngaea--ryg--ndpeffklrldwltanretlvwkaicnaw--gntyfhacn--rg- 283

```

```

OY 288 SQTNNKRCDDKGNAGKPKRAGDDVITVPTVFQVPOYLRMFEEMAEDPCKKKKKLE 347
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 284 ertkyrcrondq-----vpyfdvpyglvfwfeaaedfcrknkklk 327
OY 348 NLEKOCRGKDSDEYRYCSRNGYDCEQOTISRKGVRMGKGTDFCAFASYNWIDNQR 407
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 328 dvknrcrgkdkedrdycrsngydccktralgklyrgqscisclayacnpydwilnqke 387
OY 408 QPDKQ-KKTYKEI-----SDGGRRKRAVVG--TTKIBGYKSKSYEELKKNYGTVDNF 458
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 388 qtdkqkkkydeikkyenasgsrqkrdaggtlttnydekkfyde.lnkseyrtvdkf 447
OY 459 LGLNNKRCKDITD--GGKINFEVNSGGGVVGGSGGTSCASGNIENKCTFPRSEIC 516
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 448 leklsneictkvkdeegatidfkvn-----sdsatsagln--esgltfrskyc 498
OY 517 QPCPDGVQ--HKGG--NOMERTRVKKRWSKLYP--INGKMYLJLKSLLVYKDMMI 569
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 499 qpccpygyvkvnngsgsnweekn-gkcksgklyepkpdkegtitlilksqkghnd--- 554
OY 570 LKKMKKEFCLTNNSDGSVGYVTTGASGSEKKELDYEMKCYKINEYQKYNVQGEVDE 629
   : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 555 leeklnktodeknqgtlinsgsgtigsqgnsglqelyeewkcygedvkvghdeedee 614
OY 630 DDELKAGAGLICILPNPKKNKEVSEAKSONNHADIQKPHDFEYVWVAHMLKDSIHMRK 689
   | : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 615 dyenwnknaagliclknqknkkeeegntsekerpdeiqkltlnpflywamldslmk-k 673
OY 690 RLKSCISDQKTWKC--RNGCNKRCDFEKVWVKQKTEWPKIDHFTYCEGIPGVYFTTLE 748
   : | : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 674 klgrelqgnrlkcgknknndcecfkrlwtqkdwgjlvgfhfknlgkysgndntae 733
OY 749 LI-----LKLQFLKED-----TEBNTENSIDAEAEELKHLKIKLENNENLAV 793
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 734 lpfhdnyvlynlqeeftlkgdsedaseeksensidaeeaelkhl::eliesednqeas 793
OY 794 VNAG--TEOKTLMDKILNHELDATFKCKDCPLPEDEK----- 828
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 794 vvggyvteqnmldkllyekdeadlelehedeeekkgdgnect segenfrnypcsge 853
OY 829 ----- 828
DB 854 sgnkrypvlaankvaygmhbkaktqlasragrsalrgdislaqfkngrngstlkgqlcin 913
OY 829 -----SKGRSADPS-----PLIFTP----- 843
DB 914 enysndrsrgnsgpctlgkdgdhgyvmrlygtewsnlegkktyskrvflprprehmcen 973
OY 844 ----- 843
DB 974 lenldvsgvtnkdashslllgvqlaaktaaelikrykdnqlqldp!gqkdqeamcr 1033
OY 844 ----- 843
DB 1034 avrysfadlgdlirgrdmwdeksstmetrlltfrfknlkehhdg!kdnpyltgeskkp 1093
OY 844 ----- 843
DB 1094 aykklradwaeanhqvwramkcatkyl!cpmpvddy!pqr!rvntewaewyckagsqe 1153
OY 844 ----- 843
DB 1154 ydkllkicadsmksgdgkctqgdvdcgkkaacdyykeeleknegwtrkisdynll!ylq 1213
OY 844 ----- 843
DB 1214 aktstmgtrtvlgdddpdygmwdf!lp!bhkasi!aar!vlvraasgspteiaaap!ipy 1273
OY 844 ----- 843
DB 1274 staagy!hbelgygqceqetqfcekkgatstctkenkey!ffk!pppeyatacd!nrs 1333
OY 844 ---REBEKED----- 850

```



Dd	1334	qteepkckkeenvesacklvek1legkngrttvtgcepnkpesypdwcdkmidishnagaemp	1393
Oy	851	-----DEN-----	EDDEDEYRD-- 863
Dd	1394	prqrklclylahesqtenlktddnlkdaefiklaaaetflswyqksksndseaklldrgl	1453
Oy	864	-----DEETAKE-----TMEGS-----	875
Dd	1454	lpsqflismyrtfgdydclnclndlskkgndavakakdkgkflfsskdskspsglstrgew	1513
Oy	876	-----ADT-----	879
Dd	1514	ktngpelwkgmllcaltkyveltdcnrkikindysydkvngsqnqnpoleefaakpqflrm	1573
Oy	880	-----	879
Dd	1574	lewgeefoeerqkenllkdaenelnstqgcnadakhrcnqaeraygeyenkkkefsqgt	1633
Oy	880	-----	879
Dd	1634	nmfllkanvqgqdpheygygykdgvpqldgneyl1qkcdmnhcscmdgnvlsvspkepf	1693
Oy	880	-----TSLDVCPIYGVKYLTKDNESL0DA 903	
Dd	1694	gkyahkyepkcdcygqkhhvpsldppppvpqgqpeaetlvdvcsavl-kllfdltnfnsda	1752
Oy	904	CSLWYGCNNSRLGMRCY---TPSGEPT---SSDKRGALCVPRRRRLYIKKIYDMATKT	957
Dd	1753	cgllxy-gktapsavkcipsoelktsagatctgksdsdgsclprrrrlryvghlqwaal	1811
Oy	958	ESP0ASSEASSTSGSTTPPDSKEALLKAFVESALTEFFLHMRYKEKKAAVOAGACHG	1017
Dd	1812	-pgqegaaphsra-----dlrmaflfgaaletffldrykceekbpqg-dysqqa	1860
Oy	1018	LPRVEE-GSPVEYDEBK-LKEGKIDGFLROMFYLGRDILRSGSNDTYSVKORPS	1074
Dd	1861	lsqtlctysddeepppklllqngkldppdlrlmrylllgdyrlldlyngn--tsdsngtng	1918
Oy	1075	SSNDLKNVLLASGSTEOEREMKNRYKEI--KNFRKCSSTERSAPNLVSHPOTWENNKG	11322
Dd	1919	snm-...nlvlaesqnekdmqkldqleqllpknngtprlvpsr---aqtrpdkwnneah	19722
Oy	1133	YIMHGMICALT---SKDKLAKGVKKPKOKIENPENLMD-----ANK--PKP	1175
Dd	1973	slwgmicaltyteknptdsargde---nhkleddevyekffgstadkngtastprtyk	20299
Oy	1176	POQYTNWKLDENSGTPRTTOTQOASSDNTPTTLTHFRAPRYPFMFPEMGESFCERKK	1235
Dd	2030	tgtyevykvkledtsg-----aktpassadpr-lhsfavlrrpyfityleewgnfcfkrrkh	2083
Oy	1236	RLKQIKYDKCXYENDVC-----RCSGDGEACDSISTHNDYSTVPSFNCPCGKHCSY	1287
Dd	2084	klaqklheckveenggsstrgltitrgysdgaeacnemlphkndtvpldlekpscakrcssy	2143
Oy	1288	RKMVERKKIEFHKOSNAYGOQKTDANRNNGNPNFDKFECKTLETWPDAAKFLERLKNKGCK	1347
Dd	2144	rlwleshygkfekekakeyegqk-dkcvngsnkhnglfoetlctssakadflctl--gpc	2200
Oy	1348	TNKEYGGDDIDFEKDSKTFPHTEYGCPCPKFKTKCNCGNCGVSGJLNGCNDGDKSIDAEI	1407
Dd	2201	pnavegkrlfd---dktcfhktcdcpolkfsvnckhdecd--nskgtldcrnkhnsldatd	2256
Oy	1408	AKMSSSTTDVVMRNSDNDTTFEGDOLKQACQHANIFKSIKRDWVKCGYGVDCIDEBTN	1467
Dd	2257	engvdsfcr-lcmrysadsksngfngdglenaacraglffegldrtdekkcnvngvyvckpen	2315
Oy	1468	INERTDCKEYIOIALERPMVENFLBDYMKINDKISHCIRKKEGSKCJNGCENSKSKCEK	1527
Dd	2316	vngeaekghllqtraltvkrveyffeedynkllkhkshlrlkngelspcl-----kn--cvek	2369
Oy	1528	WIEKIAEWENIKKRFENDYENKQDPDYVNVKSTLEELPRLAIVANDQDNVTKLCFVENS	1587

Db	2370	wvdqkirkewkiterfidqykhnsoddnvrstlelllpqitdanaknkvklkfgnsc	2423
Qy	1588	GCTLTISNTQ--NNKENDAIDCMKTKLGVAKNCPGK---PSEKOSDCKEPPPLPDE---	1639
Db	2430	gcsasaneqngnqeykkaidcmklklkdkigeceekhhqtsdtesdtpqpqldeleld	2489
Qy	1640	--EDONPEENKTLPPPEFCPPPT--TQPPPEKGETCGNKKEKDKKFESEPEPAKEESGP	1694
Db	2490	dditeeaknmm--pkicenvlktqgdeqg--c-----vpaenseepatdsqk	2537
Qy	1695	AAEPPAPAESEETETNFPPEPPTGPAAPSPAP---PTPDTPLRLRPQADEPPTSL	1751
Db	2538	etpdeqpylkpkeeaavepppp-----ppqekaprpipqpppcpqpqlndpnyltai	2591
Qy	1752	QT-TRIPEGVALLAGSINFLK	1772
Db	2592	vstlswsvglgfatfctytlk	2613
RESULT 11			
W00385	ID	W00385 standard; Protein; 1726 AA.	
XX	AC	W00385;	
XX	DT	21-FEB-1997 (first entry)	
XX	DE	Truncated Plasmodium falciparum erythrocyte membrane protein.	
XX	KM	Plasmodium falciparum: erythrocyte membrane protein; malaria;	
XX	KW	detection; identification; treatment; prevention; parasite.	
OS	XX	Plasmodium falciparum MC type.	
XX	FH	Key	Location/Qualifiers
FT	FT	Domain	62..394
FT	FT	Region	/label= Duffy binding ligand domain 1
FT	FT	Domain	607..648
FT	FT	Region	/note= "Cysteine rich motif"
FT	FT	Region	839..1282
FT	FT	Region	/label= Duffy binding ligand domain 2
FT	FT	Region	1488..1523
FT	FT	Region	/note= "Cysteine rich motif"
XX	PM	W09633736-A1.	
XX	PD	31-OCT-1996.	
XX	PE	26-APR-1996; 96MO-US05798.	
XX	PR	27-APR-1995; 95US-0430908.	
XX	PA	(AFYMAX TECHNOLOGIES NV.	
XX	PI	Baruch DI, Howard RJ, Pasloske BL;	
XX	DR	WPI: 1996-497376/49.	
XX	DR	N-PSDB: T41853.	
XX	PT	New Plasmodium falciparum erythrocyte membrane proteins - used to	
XX	PT	develop products for the diagnosis, treatment or prevention of	
XX	PT	malaria parasite infections	
XX	PS	Claim 1; Figure 12; 149pp; English.	
XX	XX	A polypeptide comprising a plasmodium falciparum (Pf) erythrocyte	
CC	CC	membrane protein 1 (PfEMP1) or active fragments or analogues of that	
CC	CC	protein can be used in the treatment or prevention of symptoms of a	
CC	CC	malaria parasite infection. The polypeptides can inhibit, block or	
CC	CC	reverse the sequestration of erythrocytes in patients suffering from	
CC	CC	malaria. Nucleic acids derived from the PfEMP1 gene can be used as	
CC	CC	probes and primers to identify a Plasmodium falciparum parasite, the	
CC	CC	primers used to generate characteristic amplification patterns from	





QY	1120	VSHQITWENNNGKTYIMHGM/CALTS/DK/LAKVYKKPKIENPENLMBDANKKPPPOY	1179
Db	988	----qefwqlygkdiwkgmicalq-----aagqkltileyn	1020
QY	1180	YTNK/LDENGSTSPRTTQTOASSDNTPTLTHVKKPRYFRMEFGESFCREKRR/LKO	1239
Db	1021	ysnvtftghl-----gk/ne/asrpsllrmtwqdfcreltqlq	1065
QY	1240	IKVDCKY--ENGDVGRSCSGEACDSISTHDYSTVPFNCPCGKHSYRKMIEREKTE	1297
Db	1066	lkercmwyryngdkdygdakkek-----ctcaacyfwe/lnwqdn	1105
QY	1298	FHKOSNAYGOOK--TDATRNNGNTFDEKFCCKTLEWPDAAKFLER--LKNGPCKTKNEXGOD	1355
Db	1106	ykkngqytlewkgyspykedsdvkeaky-----ahgyrlkllhlicts-----gt	1151
QY	1356	DIDFKSKTFFQHTFYCGPPCKFKTKCQNGNGYSG/LN/CNCGKSDIAKELAKRSSTY	1415
Db	1152	d1-----ayc-----ncmeg-----fst	1165
QY	1416	DVYARVSDNDNTFEGBDLKDACQAHANIKFKIRKDWK/GYCGVDICEQTNINERTGK	1475
Db	1166	d-----sn-----	1169
QY	1476	EYIOIRALFRWENFLEDYKN/INDKISHC/KKEGSGK/CNGCKNSKCLERKWIETKIAE	1535
Db	1170	-----ndm/pselk-----	1178
QY	1536	WENIKKFPNDYOEXKQDPDYNVKSILBELP/IRAVYNDQDNWIK/LC/VEKNSGCLISNT	1555
Db	1179	-----yppl-----eleegct-----	1189
QY	1596	ONNKENDAI/CM/LK/LG/KAKNC/PKGRPSGKSDCKEPPP--LPDEQDNP/EEUTL--EP	1651
Db	1190	-----ckdpspgevlp--ekkyvpekvlpkp	1213
QY	1652	PKFCPPPTQPPPEEGGECTCGNKEKKDEKKESEEPKAEESGPAAEPPA/ASEBETN	1711
Db	1214	pkl--prrpqrker-----d	1225
QY	1712	FPEPPGTGPAAPSTYAPPPDTPPPLRQADEPFDSTLQTTIPFGA/LAGSIAFLFL	1711
Db	1226	fp-----tpa-----lknalsatimsigifratfyy	1256
QY	1772	KKKTKASVGNL/FO/LQ/PSKDYD/IDPL/LKSSNRIPYVSDR/KGKRYTYMEGSDSDKAF	1831
Db	1257	kkktksct-dllyvlnpksgdydipk/lspurylytsgkytgykrylylegsgtcs--gy	1314
QY	1832	MSDPTDVT--SSES/E/EELD/INDIYVGS/KRYKTL/LEV/L/EP/SGNWTTSAGKTP/EDTRND	1890
Db	1315	tchysdltsseseeysaelndindiyaprrpkyklllev/lepsgmnttsagmtpsdgind	1374
QY	1891	IQNDGIFSSKITD/ENWNLKKEFIS/M/LQ/N--Q/PNDYNDYTS/GNSSTNTNTTTSR/HNDV	1949
Db	1375	lqndgipskktldewm/lkldelfisqylqsgqndvpyndysqgdiprlbtq-----	1424
QY	1950	NNTWTTSR/OMMEBENLLPS/ITHG/N/LYSGE/BSY/NW/NW/NSMNDIP/IRONDVNVSGIDLI	2009
Db	1425	--pntlyfdnpekrflfisthirdlylsgseeyayvnmvntndtlpysgknglysgidll	1481
QY	2010	NDSLSGKRPID/ID/DEV/LK/RENE/LF/GEMTKR--TS/ONVAKTNTSDP/ITHNOLELHKWLD	2068
Db	1482	ndslsnsm--vdiygewl/krt/enel/fgtlnhtkhtcsnsvakelcgqdlmngdlilhkwyd	1540
QY	2069	RHRDMCEKMKNE/DI/LNK/LKEW/NKE-----NINNSGK-----TYN	2104
Db	1541	rhrdmcekmkneevldklkeewnkdnsgnlnpsgnlnpsgnlncptpsdipsgk/ldtps	1600
QY	2105	SDNKRPS/NH/VLNDV/SQ/ID/M/DNPKR/KNE/LT/NMDT/NQD/KS/TMDT/LLDLEKXNDPYRDF	2164
Db	1601	dna/lpsnktlntvcs/lq/lmdn/pkplngftn-----ndtllledlekxyneyy--d	1649

QY	2165	VEDDITF--HDVDEKSSMDITVYVHNNTSNNMDVPTKMTLENNIVNKKRIIEEETPI	2222
Db	1650	vgddltvvdhvdvstg-----snamdvpskvqtlemdl-nlk--lvkexypl	1694
QY	2223	SDIWN 2228	
Db	1695	sdwadi 1700	
		RESULT 13	
	R70235		
	ID	R70235 standard; Protein; 700 AA.	
AC	R70235:		
DT	22-SEP-1995	(first entry)	
XX	P. falciparum EBL-e2.		
DE			
XX			
KW	Erythrocyte binding ligand; EBL-e2; binding domain; malaria; therapy;		
KW	vaccine.		
OS	Plasmodium falciparum.		
XX	W09507353-A.		
XX			
PD	16-MAR-1995.		
XX			
PF	07-SEP-1994; 94WO-US10230.		
PR	10-SEP-1993; 93US-0119677.		
XX			
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.		
XX			
PI	Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;		
PI	Wellens TE.		
DR	WP1: 1995-123427/16.		
DR	N-PSDB: 083528.		
XX			
PT	New erythrocyte binding domain polypeptide(s) - isolated from		
PT	Plasmodium binding proteins, used in diagnosis, treatment and		
PT	prevention of malaria		
XX			
PS	Disclosure; Page 56-57; 81pp; English.		
XX			
CC	Erythrocyte binding ligand (EBL) family genes were cloned from		
CC	P. falciparum chromosome 7 subsegment libraries constructed during		
CC	genetic studies of the chloroquine resistance locus. The 4 genes,		
CC	EBL-e1 (083526), E31a (083527), EBL-e2 (083528) and Proj3 (083529),		
CC	encode the proteins given in R70233-36, respectively. The binding		
CC	domains of such proteins can be expressed e.g. in E. coli, yeast,		
CC	mammalian, insect, and in vaccinia virus and adenovirus-infected		
CC	cells, and provide protection against P. falciparum.		
XX			
SQ	Sequence	700 AA;	
QY	Query Match	11.4%; Score 1385; DB 16; Length 700;	
Db	Best Local Similarity	40.8%; Pred. No. 1.7e-75;	
	Matches 328; Conservative 106; Mismatches 207; Indels 162; Gaps	34;	
QY	113	CAPPRRLHCHCNKPNMNSNDSSKAKHDLAEVCAAKYEGSIRKHPKDYSKYPGSDF	172
Db	10	caprrrlhcdy---nlesidctstckllilevcaakyegsnlnhyvqhgartnedssas	66
QY	173	PWCMTLANSFADIGDILIRGRDLYLG--NKKKQNKETEREKLEQILKEIFKKIH-DNLK	229
Db	67	qlctvlarstafdigtlivrgkdllygdnkexq-----rkbleq;lkdifkllkdywmk	120
QY	230	DKEAQRKINGD-EDPNEFKYLRDQMTANRETVWAGMTSKELDNSYFRATCNDTGQGPS	288
Db	121	tnqageryyddckggdfiglredwvtsnretwkalichapkean/flkltaen-vykg--	177

OY	289	OTHHKCKDCKDKMNAKPRAGSDVTVITVTVDDYVQVLYRMFEENAEADPCKKKKKLLEN	348
Db	178	- lmgqchc-----199d-----vptyfayvpyyLwfwfeaaedfctckkkkllen	220
OY	349	LEKQCRGKDKSDERYCYSRNGYDCDQETISBKGVVRMGKCGTDCDFPACGSGYEAMINDORKO	408
Db	221	lqkgrdyeqn---lyssngydcctklykkgylvlgencstncsvcmeyetwidqkke	277
OY	409	FDKCK-KYTFEISDGG-----KKRAVSGTTYE--GYEKSEYELKNDGYGVDAFLG	460
Db	278	flkqkfyeteisgysgskprktrktraarssssddngyesklykklkwygvvdckflk	337
OY	461	LLNNEKACKOITDQCKINFEVNSGGGVGGSGGTSAGSCTDEKKGFFYSSEYQORP	520
Db	338	llnneglcqyqpyv--nekadn-----vdfnekyvktfsttelcepp	380
OY	521	DCGVQHGNGQMEKRTYVKKMRMSKLYKPIRGKMYLLLSLKVVKDMATLKKMKKEFCFLT	580
Db	381	wcgle-kggppw--kvqkgtctgsaktkttydrknitdiplyprkdsqgnllkkyhfc--	435
OY	581	ONSSDGSVGSVYTTGASGSGNSEKELYDEMKCYKHNNEQVKNVQGEVEEDDELKAGAGL	640
Db	436	-----ekgapg9gqikl--wqcyu-----dehr-----	456
OY	641	CLLPNPKNKEVSAK-----SONNHADIOKTFPHDFEYVWVAHMLKDSIHMRKRLKSCIT-	695
Db	457	--psknnnncvegtwdckltqgqct--vksyuvflfwdwvhdmlhdsavewkte-lskcin	510
OY	696	--SDGKTMKCRNKGCKDCDFEKKVVKOKETEMKPIRDHETQEGI-----PEGYV	743
Db	511	nltnngcrrnnkctcdgcfcgkvwekkqgemalikhfgkqcdlvqcgkylvlfepyg--	568
OY	744	FTTELLIK-----LQFLKEPTDEENTENSLDAEAEELKHLQILKLENNNAVYNAGT	799
Db	569	--vldlvyk9gnllqnkl-dvngdt-----ddikhlkkl--ldeedavavvlygkd	614
OY	800	QKTLMDKRLNHELNDATKCK-----DCPLPEDKSRGSDAPSPDIRPP-----	846
Db	615	ntll-dklllqhekaqaqckqkqgeckkaqesrgsaeetrederttqpadssagevee	673
OY	847	EKEDENEDDEDEVRDDEETAK	869
Db	674	eddddydeddddvqgdvavse	696
RESULT 14			
W22481	ID	W22481 standard; Protein; 700 AA.	
XX	AC	W22481;	
XX	DT	07-OCT-1997 (first entry)	
XX	DE	Plasmodium ebl-2.	
XX	KW	DBL gene family; SAMP; sialic acid binding protein; vaccine; therapy; duffy binding like gene; duffy antigen binding protein; erythrocyte; DMBP; mezozoite; malaria; var-1; var-2; var-3; var-7; immune response Plasmodium.	
XX	OS	Plasmodium falciparum.	
XX	PN	MO9640766-A2.	
XX	PD	19-DEC-1996.	
XX	PE	07-JUN-1996; 96MO-US09508.	
XX	PR	07-JUN-1995; 95US-0487826.	
XX	RA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.	

PI Chltns TE; Miller LH, Peterson DS, Slim KL, Su X;  
 PI Wellens TE;  
 XX  
 DR WP1: 1997-052231/05.  
 DR N-PSDB; T72896.  
 XX  
 XX New malaria vaccines - contains cysteine-rich DBL family protein  
 PT binding domains homologous domains of the Duffy and stialic acid  
 PT binding proteins  
 XX  
 PS Disclosure: Page 46-48; 96pp; English.  
 XX  
 CC This sequence represents ebl-2 of Plasmodium. Ebl-2 belongs to  
 CC the Duffy binding like (DBL) family of genes which have homology to the  
 CC Duffy antigen binding protein (DABP) and stialic acid binding protein  
 CC (SABP) conserved regions (see T72889 and T72888 respectively). The var  
 CC family of genes modulate cytoadherence and antigenic variation of  
 CC Plasmodium infected erythrocytes. SAMP and the Duffy antigen binding  
 CC protein (DABP) are soluble proteins that appear in the culture  
 CC supernatant after infected erythrocytes release merozoites. DABP and SABP  
 CC mediate the binding of merozoites and schizonts to the erythrocyte  
 CC surface. These proteins are necessary for erythrocyte invasion by the  
 CC parasite. This sequence can be used in the compositions of the invention.  
 CC The compositions are for the treatment and prevention of malaria, and  
 CC comprise either a nucleotide sequence or encoded polypeptide of the  
 CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of  
 CC genes having homology with conserved regions of DABP and SAMP. The  
 CC compositions are used for the treatment and prevention of malaria. They  
 CC are also used in the preparation of vaccines for inducing a protective  
 CC immune response in a mammal to Plasmodium merozoites (especially  
 CC Plasmodium falciparum or Plasmodium vivax).  
 XX  
 Sequence 700 AA;  
 XX

Query Match	11.4%	Score 1385	DB 18	Length 700
Best Local Similarity	40.8%	Pred. No. 1.7e-75		
Matches 328	Conservative 106	Mismatches 207	Indels 162	Gaps 34
OY	113	CAFFRRHLKCNKNFNPMMNSNDSSRKHBDLAEVCAAAKYEGERISIKTHPKYDSKYPGSDPF	172	
Db	10	caeyrrllhcdy---nlesitdtststkhllleevcmaakyeegnsinlthytqbrtnedeaas	66	
OY	173	PMCTYLAASFADIGDITNGRDLYLG--NKKKQKNGKETREKLEOKLKEIFKKIH-DNLK	229	
Db	67	qlctylarsfadiqdlivgkdlvlygdnkkekq-----rkkleqklxldiffkklhkdymk	120	
OY	230	DKEAQRKFRNGD-EDNPFKRLREDWMTANREYVWGMATCSKELDINSYFRATCNDGOGPS	288	
Db	121	tngageriyldackkggdfqglredwtsnrelywkallichapkeanyfllktacn-vgk--	177	
OY	289	QTHNKRCDKDKGANAGRPKAGDQVTVIPVYEDVYPOYLKWFEEBAEDFCRKKKKKLEN	348	
Db	178	-tngqchc-----iggd-----vpyfdyvpqyltwfeewaedfcrrkkkkklien	220	
OY	349	LEKQCRGADKSDERYKCCSRNGYDCQOTISRRKGVAMKGCCTDPCFACSGSYEMWIDONKQ	408	
Db	221	lqkgqrtdyegn--lycsngnydcckltylkygklvlgelchtcswcmtyewlndqkke	277	
OY	409	FDRQR-KTTELEISDDGG-----RKRRAGGTTKTYE--EYEKSEFYKLLKNDGTYDAELG	460	
Db	278	flkqgrtyetelsagsgsgskpkrktrtaarssssdnnysiekfyllkleyvgqvdskflk	337	
OY	461	LINNEKACKDITDGGKINFKEVNSGGVGVGGSGGTSASGNTDENKGTFRYSEYQPCP	520	
Db	338	lllkeglcqkqpqv--nekadn-----vdflekyvltfstrtelcepcp	380	
OY	521	DCGVQHKSGNOMERTKTKKMKRMKLYLRPINKMWLLLSKLKVVUDMMILKKNKKEFCLT	580	
Db	381	wcgle-kqgrpw--kvkqdkctcgsskctktydpknlfdlprlypdksgqnlkkyknfc--	435	
OY	581	QNSPDGSGSVYTTGASGNSSEKELYDEMCKYKHNEVQKVVNOGVEEDEDDELGAGGL	640	



• • •



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2001, 13:02:07 ; Search time 23.39 Seconds  
(without alignments) 1829.918 Million cell updates/sec

Title: US-09-508-967-1

Perfect score: 12100

Sequence: 1 MATSGGSGGTQDDAKHVLD.....YNNKKLFEFEYPISDIWN1 2228

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/CTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4817	39.8	2182	2	US-08-487-826B-16
2	4802	39.7	3060	2	US-08-487-826B-14
3	3197.5	26.4	2710	2	US-08-568-459A-12
4	3197.5	26.4	2710	2	US-08-487-826B-12
5	1385	11.4	700	2	US-08-568-459A-10
6	1385	11.4	700	2	US-08-487-826B-10
7	702	5.8	921	2	US-08-568-459A-8
8	702	5.8	921	2	US-08-487-826B-8
9	418	3.5	1435	2	US-08-568-459A-4
10	418	3.5	1435	2	US-08-487-826B-4
11	379.5	3.1	362	2	US-08-568-459A-18
12	379.5	3.1	362	2	US-08-487-826B-30
13	360.5	3.0	411	2	US-08-568-459A-19
14	360.5	3.0	411	2	US-08-487-826B-31
15	345.5	2.9	749	2	US-08-568-459A-6
16	345.5	2.9	749	2	US-08-487-826B-6
17	315	2.6	1115	2	US-08-568-459A-2
18	315	2.6	1115	2	US-08-487-826B-2
19	315	2.6	1115	6	5198347-6
20	311.5	2.6	1663	5	PCT-US93-07261-16
21	288	2.4	1588	5	PCT-US93-07261-11
22	262.5	2.2	3969	4	US-08-061-376-5
23	226.5	1.9	1183	2	US-08-447-031A-2
24	220	1.8	2843	1	US-07-741-940-2
25	220	1.8	2843	1	US-08-289-548A-2
26	220	1.8	2843	1	US-08-452-654-2
27	220	1.8	2843	2	US-08-370-235A-2

28	217	1.8	1706	2	US-08-459-568-2	Sequence 2, Appl1
29	217	1.8	1706	2	US-08-399-411-2	Sequence 2, Appl1
30	217	1.8	1706	3	US-08-516-859A-2	Sequence 2, Appl1
31	217	1.8	2842	1	US-07-741-940-7	Sequence 2, Appl1
32	217	1.8	2842	1	US-08-289-548A-7	Sequence 7, Appl1
33	217	1.8	2842	1	US-08-452-654-7	Sequence 7, Appl1
34	217	1.8	2843	1	US-08-452-655B-2	Sequence 2, Appl1
35	217	1.8	2843	1	US-08-452-655B-7	Sequence 2, Appl1
36	217	1.8	2843	4	US-08-450-582-2	Sequence 2, Appl1
37	217	1.8	2843	4	US-08-450-582-7	Sequence 7, Appl1
38	217	1.8	2973	2	US-08-821-355A-7	Sequence 7, Appl1
39	217	1.8	2973	2	US-09-003-687A-7	Sequence 7, Appl1
40	217	1.8	2973	4	US-09-136-605-7	Sequence 7, Appl1
41	216.5	1.8	1719	2	US-08-459-568-4	Sequence 7, Appl1
42	216.5	1.8	1719	2	US-08-399-411-4	Sequence 4, Appl1
43	216.5	1.8	1719	3	US-08-516-859A-4	Sequence 4, Appl1
44	213.5	1.8	2353	4	US-09-377-155-33	Sequence 33, Appl1
45	213.5	1.8	2353	4	US-08-913-942-4	Sequence 4, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-487-826B-16  
Sequence 16, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chlntis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESS: Knobb Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487, 826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29, 655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2182 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
US-08-487-826B-16

Query Match	39.8%	Score 4017	DB 2	Length 2182	
Best Local Similarity	45.9%	Pred. No. 0			
Matches 1096	Conservative 298	Mismatches 607	Indels 388	Gaps 93	
QY	4	SGGSG---GTOD-----	EDAKHYLDERGQVHNE--VHGAKNVYSLKSGSLIAS-I	50	
DB	11	SGGSSGGKKKKDTSIYYSDAKDLDRYGEYEEKAYANGAKKTYEALKGNLTANR	70		
QY	51	LGETAFTVAKSMOTESKYTELIEANSKRNKCKDKGNDREPS-----	-YKEQAG	99	
DB	71	SSETASISIECTLVKYEYERVANGDGRHRCRDAKNEVDNRRSDTLGGGCTYNRKIDSQO	130		
QY	100	YDNKKKKCSNGMTCAPFRLLHLCKNFPNMSNDSKAKHDLLAEYCAAKYEGESIKTH	159		
DB	131	GNKV-----GACAPYRLHLCDY---NLESIDTSTYHKLLLECYMAKKEGSIINTH	181		
QY	160	YPKYDSKYPGSDPEMCTMLARSFADIGDIRDRDLYLG--NKKKKONGEYREKLEQTL	217		
DB	182	YTOHOTNEDSASOLCTYLARSFADIGDIRKDYLDGDNKEKO-----RKKLEQTL	235		
QY	218	KELEFKIIR-DNLKDEAKQRYNGD--EDPNEYLRREDMWNRETYWAMTCSKEKIDNSY	275		
DB	236	KDIFKKIHDVAKTNGAQRIRYIDDAKAGDFQLREDWMWTSNNEYWKALICHAPKEANVF	295		
QY	276	FRATCDTQGSQSHNKKCRCDKKGANAKFRAGDGVTVPTFYFDVYPOYLRMPEMA	335		
DB	296	IKYACN-VERK---TNGQCHC-----IGGD-----VPTYFDVYPOYLRMPEMA	335		
QY	336	EDPCRRKKKKLEMLEKQCGKDKSDERYCSHNGYDCBOTISRKKCVBMGKCTCCTCFPC	395		
DB	336	EDPCRRKKKKLEMLEKQCHDYEN---LYCSGNGDCTTYTRKKGVLVGEHCTGCSWC	392		
QY	396	GSYEMNIDNQKQFODKQ-KYTRKELSDGG-----RKRKAVGTTKYE--GYEKSFEYKL	447		
DB	393	RMVETWIDNQKKEFLQKRYELETISGGSGSPKRTKRAANSSSDNGYENSFYKTL	452		
QY	448	KNDYTVAVAFGLNLNNEKACDITDGKINREKVNSSGGGVGGSGGTSGASGTNDENK	507		
DB	453	KEVGYODVDFLKLKEBICQKOPQV--NEKADN-----VDFTNEKYV	495		
QY	508	GTFFREYECOPDQGVQKGNOMERKYYKMRKSKLYRINKMVLKLSLVVDM	567		
DB	496	KTFESRLEICEPCWCOL-KGGRPM--KKAGDTGCSAKTKYDPDNITDIPVLPDSQ	552		
QY	568	MILKNMKEFECTLONSSDQSVSVYTTGASGENSEKELDYEMKCYKHNQVQKVVQGEV	627		
DB	553	ONILKKYKNFC-----EKAGPAGGQIHK-----WQCY-----	580		
QY	628	EBDDDELKAGGLCTLPNKKKEVSEAK-----SONNHADIOKTHDFEYVVAHMLKDS	683		
DB	581	---DEHR-----PSSKNNNCCGEGTWDFTOGQT--VKYSNVFPMVDVHMLHDS	626		
QY	684	IHMRTKRLKSCI---SDGKTMKCRNGCNKKKCCCFEYWKQKETEKKPIKDNKKYDEGT--	738		
DB	627	VEKTE-LSKCTINNNTNGMTCRNNKCKTDCGCFKQWEKKQOEWMALKDHNKGTQDIYQ	685		
QY	739	-----BEGYFTTLELILK-----LOFLKEDTEBENTENSIDAEAEELKILQILKLE	786		
DB	686	QKGLVFSYSG---VLDVLKGNLQAIK-DVINGDT-----DDIKIKKL--LD	729		
QY	787	NENNLAVNAGTEQKTLMDKLLNHELINDATYCK---DCPLREEDKSIRGSADSPDIFI	842		
DB	730	EBDAVAVVLGGKDNPTI-DKLQNEKEQAEQCKQOECEKKAQOESRORSARETREDERT	788		
QY	843	PPP-----EKEEDENEDDDDEYAVADEETAKETTESGATDTTISLIV-----	885		
DB	789	QOPADASAGEVEEBEDDDDDVDEDEDDVQEEBEKE--EGTVETVTEVTEVEETVTEQ	846		
QY	886	-----CPYKGLVTLDNESLQDASGLY--GGNNSRLGRCVTPSGPEPTSDKGAICV	938		
DB	847	EGVKECODIVGK-LFEDQSKLEACGLKIGPQGEKEFPNNKCYTPGCVSTATSGKGAICV	905		
QY	939	PPRRRLVYTKKIVDWATKTESPOASGSEASSTSGSTTPDCK---EALLKAFVEEALET	995		

Db	906	PRRRRLVVGGLSQWASR-----GGDETEVVSSEATISAPSOSESERKLTAETAEALET	959
Qy	996	PELHMRYEEKKAAVA-OBGAGHGLPRVEEGSPEDDEKLKE-GKIPDGLQWMEYTLGD	1053
Db	960	PELHMRYEEKKPRATODGAGLGSLEBPSPREDDPQTOLOQGVIPREFLQWMTYTLAD	1019
Qy	1054	YRDILEGSSNDT--SVSKDTPSSNDMLKNI VILLASGSTEOREKKN---KYKEIKNFR	1108
Db	1020	YKDILYSSNDTSPTTGKQTPSSNDMLKNI VLEASGSTEOEKKEMQIOAKIKTLNGA	1079
Qy	1109	KCSTRSAPNLVSNP-QTWENNNKYTWBGVCAVLSKOKIAKVCK-PQKILENPEML	1166
Db	1080	TSGVPVTKNSVKNPQOTWMENTAKLDIMNAMVCAVLYKKENDAGTSAKLEQKKDLKMLM	1139
Qy	1167	DEANKPKPRPOLYOTNYKLDENSTSPRTQTOASSDNT--PRTTLHFVRYPRTRWEE	1224
Db	1140	DEANKNTIEKYQTTNKLDESG-----AKSNDTIQPTTLKNFEITPFRMLHE	1190
Qy	1225	WGSEFCREKRRLKLOIKVDCVENGDVRCSGDGEACDSTIHDYSTVPSFNCPCGKHC	1284
Db	1191	WGNSECFERAKRLQIKHECDEDEG-KQXSGGECEIFSKQYVNLQDLS-SSCARKC	1248
Qy	1285	SSYRWIERKRLIEHKOSNATGOQKT-----ATRNGTPEKCEKCTLEWPPAA	1333
Db	1249	RLYKWIEMKKTLEYEKQOAKAEQOKSVYENEOKDQOTOSNNNA--NEFSRLGASPPAA	1306
Qy	1336	KPLEELKNGPCKTKNEY-GGDD--IDPEKSDKFFOHTCEKCPCKFKTKCQNGNCGVSGL	1392
Db	1307	EPLQLT--GSKKNDNGENGEDNKIDPKNDKTFKKAHSDCEPRTIGVACQNGHC-VGSA	1366
Qy	1393	NG-NCDGKSIDAKEIAKMRSSTTDVAVRVSNDJNTFE--GDCLKDACOHANIEGIRK	1449
Db	1364	NKECKKNK--ITAEIDKKTDPNCINIMVAVSDJNTFEHLD-----CKSGSIEGIRK	1417
Qy	1450	DYMKGYCVGDI--EQNTINETDCKEIQIRALFKRVNFELEDDYKINDKISHCIT	1507
Db	1418	DEMKCANVCGVDTLEKKNKGEGKKYITKELKRLFELEDDYRNIRKIKLCTK	1477
Qy	1508	KGEKSGINGCEKNSKCLKEIKIEWENIKRRFNDQYENKDQDQVWKSILEELPK	1567
Db	1478	KEDGCKCIKG-----CLEKWQDCKTEMKINDYTLBQKND-----GNTLNFLEQ	1522
Qy	1568	IAVVDODNVTKLC---VENSKCTLISNTQNNKENDAIDCMKLIGVAKANCGRPS	1623
Db	1526	FQYRTFENKALPCDGDLDQFTSCGLNSTQNGNNNDVLCTLANKLOKISECKEHS	1585
Qy	1624	GKQSDC-----KEP---PRLDEBDONPEENTLEPRFKCPPTTQPREEKGGE--TGC	1671
Db	1586	GOTQPCPNSSLSGKESTLEVEDVDYDEQNP-ENKYEOPFCFCDMKPEKKNDEEGTGC	1644
Qy	1672	NKEEKDKKEESEPAKKEESGPAKEEPAP--FAESEFETN-FRPECTGPAAPSPRA	1728
Db	1645	GDEKK--KVEDSVIEQKEEBAASAPESPRLPEAPKKEENVKRP-----	1690
Qy	1729	PPTPTPTPT-----LRPADER-FDSTIIQTTIPRGVALALGSLAFLKLLKKTASVGN	1781
Db	1691	-----PPPKRRRTKTRNVLDHRAVIRPALMSSTIMMSIGIGFAAFYFLYLLKKTSSVGN	1744
Qy	1782	LFQILQIPKSDYDIPFLKSSNRYIPVSDRYKCKTYIYEGSD-EDKAYFASDTTQVS	1840
Db	1745	LFQILQIPKSDYDIPFLKSSNRYIPVSDRYKCKTYIYEGSDSGDEKAYFASDTTQVS	1804
Qy	1841	SESEYEELDINDIYVPSPKYLLIEVULLEPSSGNNTAGSKMPTSDTRNDIOND---GIP	1897
Db	1805	SESEYEELDINDIYVPSPKYLLIEVULLEPSSGNNTAGSKMPTSDTRNDIOND---GIP	1853
Qy	1898	SS-----KITDENMQLKEFISNMLQNPVDYNDTSGNSSSTNTNITTSRHNVDNNTN	1953
Db	1854	NSDTPPTTDDENMQLKDDTISNMLQNTONTPEH-----ILHDVNDNTH	1898
Qy	1954	TTMSRDNNEENLLPSIHGNTLSGEEYSYNV-----NMVN-----SMNDIPI	1966

Db 1899 PTMSRHMNDQKPFMSIHDRNLFSGEERYNDMFNSGNPNINISDSTNSMSTLNNHSPY 1958  
 QY 1997 NRDNNVSGIDLINDISGCKPIDYDEVLKRRKENLFGTEMTKRTSTON--VAKTNSD 2054  
 Db 1999 NKNLDYSISDIDLINALSGNH--IDYDEMLKRRKENLFGTQHPKRTITSNRVYOTSSD 2017  
 QY 2055 PIHNOLELFHKMLDRHRCCEKMKNEEDILNKLEEMKENENNSGKTYSNPKPSHNV 2114  
 Db 2018 PTTNOINLPHKLDHRHRCCEKMKNNHRLPKLELM--ENETHSGDT--NSGIRSGNHV 2073  
 QY 2115 LMTDVSIGIDMNPRTKNETMTNODKSTMDTILDLLEKYNDRPYDYEDDIYHDV 2174  
 Db 2074 LMTDVSIGIDMNPRTKNETMTNODKSTMDTILDLLEKYNDRPYDYEDDIYHDV 2132  
 QY 2175 DYKSSMDIYDHNNTYSNNMDVPTKMIEMNIYNNKKEITEEPIS 2223  
 Db 2133 NDKKASEDHINMDHNMNNNSDPTNVOIEMNINN--OELLONEYPIS 2180

RESULT 2

US-08-487-826B-14  
 ; Sequence 14, Application US/08487826B  
 ; Patent No. 5993827  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sam, Kim L.  
 ; APPLICANT: Chitnis, Chetan  
 ; APPLICANT: Miller, Louis H.  
 ; APPLICANT: Peterson, David S.  
 ; APPLICANT: Su, Xin-zhaun  
 ; APPLICANT: Wellens, Thomas E.  
 ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
 ; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
 ; NUMBER OF SEQUENCES: 45  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe Martens Olson & Bear  
 ; STREET: 620 Newport Center Drive 16th Floor  
 ; CITY: Newport Beach  
 ; STATE: California  
 ; COUNTRY: US  
 ; ZIP: 92660  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/487,826B  
 ; FILING DATE: 10-SEP-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Israelien, Ned  
 ; REGISTRATION NUMBER: 29,655  
 ; REFERENCE/DOCKET NUMBER: NIH121.001CP1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619) 235-8550  
 ; TELEFAX: (619) 235-0176  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3060 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-487-826B-14

Query Match 39.7%; Score 4802; DB 2; Length 3060;

Best Local Similarity 36.2%; Pred. No. 0; Matches 1156; Conservative 277; Mismatches 631; Indels 1128; Gaps 83;

QY 4 SGGSGTODEDEKAKVLDGFGQVHDEVEHGEAKNYVELSGSLASLILGETAFTVSMQT 63  
 Db 12 AAGGDIDIESAKHMFDRIGKDYVDKVKREAKERKGLGRLS-----EAFKEKESNP 65

QY 64 ES-----KYTELIEANSKRNPCK-----KDKGNDVDFRSYRBOAGYDNKMKK 107  
 Db 66 QTPEDPCDLHDKYHTVNTN--VINPCADSDYRFSDEYGCOTHNRIKISQGDNGK---- 121  
 QY 108 SNGMTCAPRRLHLCKNKPNNNSDSKAKHDLAELVCAKREGEKTKTHYPKYDXY 167  
 Db 122 ----ACAPYRLHYCDQNLQIEPIKITNT--HNLVDVCAKAFBEGQITQDPKQOAY 176  
 QY 168 PGSDPFCMTLARSPADIDIDIRGDLVLYGNKKKQNGKETEREKLEK/LKLEIFKKIDN 227  
 Db 177 GDSPOICTMLARSPADIDIVGRDLYLGNPOEIK-----OROOLEINLKITFKIYER 231  
 QY 228 LKDKERQKRYNDEDDPNFKLREDMTARERVWGMGMSKELDNS--FRACNDNGOP 287  
 Db 232 LGAELA--RYG--NDEFFKLEDDMTARERVWKAITCNAM--GNT/THATCN--RG- 281  
 QY 288 SOTHNRCRCDKDGANAGKPKAGDGVTVPTVFDVVPQYLWFEEMVEDFCRKKKKLE 347  
 Db 282 ERTKGYCRNDQ-----VPTVFDVVPQYLWFEEMVEDFCRKKKKKKIK 325  
 QY 348 NLEKQCRGKDSDEYRYCSRNGYDCBOTISRKQVYMGCTDCFPFAJGSTEWNIDNQR 407  
 Db 326 DYKRCRGKDKEDKDRYCSRNGYDCBETKRAIGKLYRGQCSICLYACNRYDWMNQKE 385  
 QY 408 QFDKQ-KKYTEI-----SDGGKRKRAVCG--TTKYGEYKSFYEKAKNDGYTVADF 458  
 Db 386 QFDKQKKYDEIKKYENGASGSRKRDAGGTTNTYNGEKFETELINKSEYRVDFK 445  
 QY 459 LGLNNEKACKDITD--GSKINPKEVNSGGGVGGSGGSGTSGASGTDNNKGFYFSEVC 516  
 Db 446 LEKLSNEEICTYVKDEEGTIDPKNKN-----SDTSGASGTTNVSOGTFFYKSKC 496  
 QY 517 QPCPDGVQ--HKG--NOMERTKYKMKWSKLXKP---INGKMYLLKSLKVRKDM 569  
 Db 497 QPCPYGVKVVNNGSSNMEKKN--GKCKSGKLYPRKDKGTTTLLKSGGHND--- 552  
 QY 570 LKKNKEFCLTQSSDSGVSQVYTTGASGNSKELLYDEMKCYKHNQKVVNGVEVE 629  
 Db 553 IEKLNKFCDEKNRGDTINGSGGTGGGNGSRQELYEEMCYKEDVYKGDHDEDE 612  
 QY 630 DDELKAGAGLCILPNPKNKEVSEAKSONNHADIDKTFHDFYFY/AMHLDKSHMPTK 689  
 Db 613 DYENKAGAGLCILNKKNNKEBEGGNTSEKPEIDKTTNPFYTYVAMHLDKSHMK-K 671  
 QY 690 RLKSCISDKTYMC--RNGCNKCDCEKVVQKETEEMKPIDKHFTJDEGIPRGYFTLE 748  
 Db 672 KLQRCLONGNRJIKCGNNKCNNDCEFKRWITQKDEMGKIYQHFQNIKRGSGDNTAE 731  
 QY 749 LI-----LKLQFLKED-----TEENTENSUDAFAEBELKHLQKILKENENLAV 793  
 Db 732 LIPFDHYLYOYLQEFLEKGSDESASEKSEMSDAEAEELKHLREIIESBDNQEAS 791  
 QY 794 VVAG--TEOKTMDKLNHLELDATKCDPLPEDEK----- 828  
 Db 792 VGGGTQKNIHDKLNTYKEDRADLCLEIHEDEBEKEKGDGNECIJEBGENTRYNCSGE 851  
 QY 829 ----- 828  
 Db 852 SGNKRYVLANKVAYQMHKAKTQLASRAGSALRGDISLAQFKMIRNGSTLKGQICKIN 911  
 QY 829 -----SRGRSADPS-----PIIFIP----- 843  
 Db 912 ENYSNDSRGNSGPCYTKGDKDGGVBMRIGTSEWNIIEGKQYKQVFLPRPREHMTSN 971  
 QY 844 ----- 843  
 Db 972 LENDVGSYTKNDKASHSLGVDYQLAKTDAEIIKRYDQNNIQUTDPIDQKDEAMCR 1031  
 QY 844 ----- 843  
 Db 1032 AVRYSPADIGDILRGDMDEDEKSSITDMETRLITVFNKIKKHDGKIDNPKYTGDESKKP 1091

0Y 844 ----- 843  
Db 1092 AYKKLRADMEANRHOVRAAMKCATKGIICGMPYVDYIPIORLRMTWMAEMCKAOSOE 1151  
0Y 844 ----- 843  
Db 1152 YDLKKICADCMKSGDKCTOGDVDCGKACADCKYKEEIEKMNQWRKISDKYNLYLQ 1211  
0Y 844 ----- 843  
Db 1212 AKTSTNGRTVLGDDDDPYQOMVDFLPIHKASTAARVLVKRAAGSPTEIAAAPTTPY 1271  
0Y 844 ----- 843  
Db 1272 STAAGYIHOEIGYGCOEOTQFCEKKHGATSTSTTKENKEYTFKOPPEYATACDINRS 1331  
0Y 844 ---PPEEKD----- 850  
Db 1332 QTEBPKKKEENVESACKIVERILEGKNGRTTVGECNPKESYPMDCNNIDISHGACMP 1391  
0Y 851 ----- 863  
Db 1392 PRROKLCIYTAHESQTFENITDNLDAFTKTAAEFLSMWYTKSKNDSEAKILDNGL 1451  
0Y 864 ----- 875  
Db 1452 IPSOFLSMWYTFGDYRDICLNTDISKONDVAKAKDKIGKFKSGDKSPSGLSROEMW 1511  
0Y 876 ----- 879  
Db 1512 KTNGPELWKGMLCALTYKYDTDTDNKRRIKNDYSDYKVNQSGNPSLEFAKPOFLRMW 1571  
0Y 880 ----- 879  
Db 1572 IEMBEFCAROKKENIITKACNEINSTQOCNDAKHRQNCACRAYOEYVENKKKEFSQT 1631  
0Y 880 ----- 879  
Db 1632 NNFLKAVNOPODEYKGYEKDGVPIQNEVYLQKCDNNKSCOMDGNVLSVSPKEKF 1691  
0Y 880 ----- 903  
Db 1692 GKVAHKYPEKDCYQGHVPSIPPPPPVOPQPEAPRTVYVCSIV-KTEKDKDNNESDA 1750  
0Y 904 CSLKYGNNSRLGRVY---TPSEPT---SSDKNGALCVPPRRRLYIKKIYDMATKT 957  
Db 1751 CGLKY-GKTAASSMKCIPSDTKSGAGATGKSGDSGSICIPRRRLYVYKLDWMATL 1809  
0Y 958 ESPQASGEASSTSGSTTPPDSKELALKAFVESALETFLMHRKKEKKAVAOEGAGHG 1017  
Db 1810 --PGEGAPSHSRA-----DRLNAPFQSAAIETFFLMDRYKKEKKPQG-DGSGQA 1858  
0Y 1018 LPRABE--GSEYEPDEK-LKEGKIPOGFLQMFYTLADYDILFSGSNDPTSVSKDTPS 1074  
Db 1859 LSOJLSTYSDDEBPPKLLONGKIPPEFLMLEYTLGDYDILVHGN--TSOSGNTNG 1916  
0Y 1075 SSNDNLKIVLLASGSTEOREKKNKYKEI--KNEKRCSTERSAPNLVSHQPTWENNKG 1132  
Db 1917 SNNN---NIYELASGNKEDMKIOEKIEQILPKNGGTPVLYKSS---AQPTDKMNEHAE 1970  
0Y 1133 YIMHWGVALT---SKDKIAKVEKKPKOKIENPENLME-----ANK-----PKP 1175  
Db 1971 SIMGMICALTYTEKNPDTSARGDE--NKTEKDEYERKEFFGSTAKHGASLPTGYK 2027  
0Y 1176 POYOTYVAKLDENSGTSPRTTOTQASSDNPTTLTHFYKRPYTRWFEWGESCREBK 1235  
Db 2028 TOYDYEVKLEDTISG---AKTPSASSDTP-LLSDFLRPPRYLYEWMQNCCKRRKH 2081  
0Y 1236 RLKQIKVDCYKENDVG-----RCSGOEACDSISTHDYSTVSPNCGCGKHCSY 1287  
Db 2082 KLAQIKHECKVEENGSGSRKGITROYSGDEACNEMLPKMDGTVPLDEKSKPCSSY 2141  
0Y 1288 RWTIERKKIEFHKOSNAYGOOKTDATRNNGNTFDKEFCLETWPPDAKFLERLKNPCK 1347

Db 2142 RMTIESKKEFEKQEKAEQOK-DCVNGSNKHNGCELTLTSSAKDFLKL--CPCK 2198  
0Y 1348 TNKEYGDDIDPEKDSKTFQTEYCGPCPKRTKQONCGVSLNCGODKSIDAKEI 1407  
Db 2199 PNNVEGKTIFD--DGRFTKTKDCCPLKFSVCKKDECD-NSKGTDCRRKNKSIDATDI 2254  
0Y 1408 AKMRSSTTDVVMARYSDNDNTNFEGBDLKDACOHANIFKGIKRDYKMGCGYGVADICEQTN 1467  
Db 2255 ENGVDSIV-LEMRYASBSKSGFNGDGLNACRAGIEGIFKDKEMKCRNNGYVVCAPEN 2313  
0Y 1468 INERTGKEYIQRALFKRWYENFLEDYKINDISHCIRKBEKSGKINCENKSKLEK 1527  
Db 2314 VNGEAKGKHIIQIRALYKRWYEFEDYKIKHRIKSHIRKNGEISPCI---KN--CWEK 2367  
0Y 1528 WIEKKIAMENIKKRPNDQYNNKQPOYVWKSILBELPKIAYVNDQNVKILCVFENSK 1587  
Db 2368 WVDOKRKEWKEITERFDQYNDNSDDNVSFLETILPQITDANAKNKYIKLSKFSNGC 2427  
0Y 1588 GCTLISNTQ--NNKENDAIDMLKKGKAKNCPGK--PSGEKSDCKEPPPLPDE--- 1639  
Db 2428 GCSASANDQNKNGEYKDAIDCMLKKLKDIGCEKKNHQTSDTSCSDTPQOPQTELETL 2487  
0Y 1640 ---EDQNPENTLEPPKFCPT--TPPEKGGETCGNKEEKDEKKEESEPAKESGP 1694  
Db 2488 DDIETEAKKNMM--PKICEVNLKTAQOEDEGC---C-----VPAENSEEPATDSGK 2535  
0Y 1695 AAEPAPTAESEETETNPPRPGTGPAPSTAP---PPDPPTPLRPQADEPFDSTIL 1751  
Db 2536 ETPEOTPVLPKEEBAVPPPP-----PPOEKAPAPQOPPTPQTQILDNHVLTAL 2599  
0Y 1752 QT-TIPGVALALGSIAPLFLKRTKASVGNLPQILOIPKSDYDIPRLKSSNRYIPVSD 1810  
Db 2590 VSTLANSVGIQFATPFYFYIAKKTSSVGNLPQILOIPKSDYDIPRLKSNRRIPTSG 2649  
0Y 1811 RYKGRYIYMEGSDDEKRYAASDTPDVTSSESEYEELINDIYVPSPKYTLILEVLE 1870  
Db 2650 KYRGRYIYEGSDGJDS-GYTDHYSIDTSESEYEEMDINDIYVPSPKYTLILEVLE 2708  
0Y 1871 P-----SGNNTASGRNTPSOFRNIOINDGIPSSKITNENWQLKKEFTSN 1916  
Db 2709 PSGNNTASGNNTASGNNTASGNTPSDIONDIONGIPSSKITNENWQLKDEFISQ 2768  
0Y 1917 MLONOPNDPNDYTSNGNSTNTITTSRHNVDNNTNTMSROMMEENLLPSIHDNLV 1976  
Db 2769 YLOSEPNTERN-----MIGYVNDKNTPTSHHANEVEKPTFMSIHDNLF 2813  
0Y 1977 SGEYSYNV---NMVN-----SMNDIPINRDNNVSGIDLINDLSGKPI 2019  
Db 2814 SGEENYDMFSGNPNPISIDSTNSMDSLTSNHSPYNDKNDLGSIDLINDALSNGH-I 2872  
0Y 2020 DIYDEVLRKRENELEFTE-NKRTSTONVAKTMSDPIHNLLEFHNWLDHRMCEKWK 2078  
Db 2873 DIYDEMUKRKRENELEFGKHNHKAHTNTYNAKPARDDPTQOINLFHNWLDHRMCEKWK 2932  
0Y 2079 NKEDILNKLKEEMKENINNSGKTYNSDNKSHNVANTVDSIOIOMDNPXTKKEIENMD 2138  
Db 2933 NNHERLEKRLKELM--ENETHSGDI--NSGIPSGHNVLNTVDSIOIOMDNPXTKKEIENMD 2988  
0Y 2139 PTMKNHIEKNIVN 2210  
Db 3049 PTMKNHIEKNIVN 3060  
RESULT 3  
US-08-568-459A-12  
; Sequence 12, Application US/08568459A  
; Patent No. 5849306  
; GENERAL INFORMATION:



```

Db 1394 PRROGLCLYIAHESQTEINIKTDNLKDAFTKTAAEFTLSMYYKSKNSDEAKTILDRGL 1453
Qy 864 -----DEBTAK-----TTEGS----- 875
Db 1454 IPSQFLRSMATTFGDYRICLNTDISKONDYAKAKDKIGAFKSGDKSGSLSRQEMW 1513
Qy 876 -----ATDT----- 879
Db 1514 KTNPEIWKGMICALTKYVTDNKRKIKNDYSYDKVQNGNPNLSLEPAKQFLRM 1573
Qy 880 ----- 879
Db 1574 IEMGEFCAEROKKENIKDCAINEINSTOQCNDAKHRCNACRAYOEYVENKKKEFGSGOT 1633
Qy 880 ----- 879
Db 1634 NNFVLKANVOPQDEPYKGYEKDGYOPIQGENEYLLOKCDNNKCSOMGNVLSVSPKEKPF 1693
Qy 880 -----TSLDVCPIYGVLTAKDNESLQDA 903
Db 1694 GKYAHKYPEKCDYQGRKHVSIPEPPPPVQPPPEAPITVVOCSIV-KTLKDTNNFSDA 1752
Qy 904 CSLTYGCGNNSRLGRV---TPSGEPTT---SSDKNGAICVPPRRRLYIKKIYDWTXT 957
Db 1753 CGLYV-GKTAPSSMKCIPSDTKSGCATTKSGSDSGSICIPRRRLRYVGLQEMATL 1811
Qy 958 ESPASGSEASTSGSTTPPSKELALKAFESAAIEFFELMHYKKEKKAVADEGAGHG 1017
Db 1812 --PGEGAAPEHSHRA-----DRLRNFQSAIETFFLMDRKEKKKQOG-DGSOQA 1860
Qy 1018 LPRVEE-GSPEDYDEDK-LKEGKIIPGFLRMEYTLGDYBILFSGSNDTTSYKSDTPS 1074
Db 1861 LSQLTSYSDDEEDPPKLLNGKILPPFLRLMFTYTLGDYDLVHGN--TDSQNTNG 1918
Qy 1075 SSNDNLKNIYVLASGSTEQERKNNKYEI--KNFRKCTERSAPNLVSHPOTWENNGK 1132
Db 1919 SNNN--NIVLEASGNKEDMOKIOEKIEQILPKNGGTPLYVERKS--AQTPDKMNEHAE 1972
Qy 1133 YIMHGWALC-----SKDKIAKGYEKKRQKLEPENLMD-----ANKK-----PKP 1175
Db 1973 SIMGMICALYTEKENDPDSARGDE--NTEKDEVEYEKFGSGTADHAGSTAPPTGYK 2029
Qy 1176 POYOYTNKLDENSGTSPTTQTOASSDNPTTLTHFYKRPYFRMEEGESFCREKREK 1235
Db 2030 TOYDYKWKLEDTSG-----AKTPSASDTP-LSDPFLRPPRYRILEBMQNCCKRKH 2083
Qy 1236 RLAKIKYDKVKGENDVG-----RCSGDGEACDSISTHDYSTVPSPFCGCGKHCSSY 1287
Db 2084 KLAQIKHECKVEENGSGSRGKITROYSGDEACNEMLPKNDGTVPLDEKPSCAKPCSSY 2143
Qy 1288 RKWIERKKIEFHKOSNMYGOOKTDATRNNGTPEKFCQKTELETPMDAKELERLKNCK 1347
Db 2144 RKWIESGKEKEKOEKAYEOK-DKCVNGSKKHNGFCETILTSSKADFKTL--GPKC 2200
Qy 1348 TNKEYGDDIDFEKDSKTFQTEYCGPCPKFKTNCQNGCNGVSGLNGCDKSDIAKEI 1407
Db 2201 PNNEGKTIIPD--DKTFKTKDCDCLKRSVCKKDECD-NSKGTDCRKNKNIDATDI 2256
Qy 1408 AKMRSSTTDVVMARYSDNDJNTFEGBDLKDACQAHNIRKIGIKYDKWKGCGYCVDICEQTN 1467
Db 2257 ENGVDSIV-LEMRVSAKSGFNGDGLNMACRGAGIEGIRKDKMCKGCMVCGYVYCKREN 2315
Qy 1468 INERTDGKEYIQIRALPKRWENFLEDYNNKINDKISHCIRKGEKSGKCNCEKSKCKEK 1527
Db 2316 VNGAKAKKHIIQIRALPKRWYEFEDYNNKIKHISHRIKNGELSPCL---KN--CVER 2369
Qy 1528 WIEKIAEWENIKRFRFYDQYENKQDQPYNKSILEELIPKIAVVDNDQNTIKLCVFENSK 1587
Db 2370 WVDQKREMKRITERFKDYKNDNSDDDNVRSFLETILPQITDANANKVYIKLKFQSGC 2429
Qy 1588 GCTLISTNQ--NKNENDADCMKLKLVKAKNCPGK---PSGEGOSDCKEPPPLPDE--- 1639

```

```

Db 2430 GCSASANEQNNKNGEYKDAIDCMKLKLDKIGECSEKHHQTSDECSPTPOQOTLEDTLD 2489
Qy 1640 ---EDQNEENTLEPPKCPPT--TOPPEKGGFTCGNKKEKDEKKESEEPAKESGP 1694
Db 2490 DDIETEAKKMM--PKICEVNLKTAQOEDBG---C-----VPAENSEEPAATBSGK 2537
Qy 1695 AAEPAPTAESEETETNEPPEPGTGPAAPSTPAP---PTPDPTPLRPOADEPFDSTIL 1751
Db 2538 ETPQOTPYLKPEEVAVEPEPPP-----PPOEKAPAPIQPPPTPTQLLDNPHVLTAL 2591
Qy 1752 QT-TIPGVALALGSIAPFLK 1772
Db 2592 VTSTLWMSVGIGFATFTFYFLK 2613

```

## RESULT 4

```

US-08-487-826B-12
: Sequence 12, Application US/08487826B
: Patent No. 5993827
: GENERAL INFORMATION:
: APPLICANT: Sim, Kim L.
: APPLICANT: Chitnis, Chetan
: APPLICANT: Miller, Louis H.
: APPLICANT: Peterson, David S.
: APPLICANT: Su, Xin-zhaun
: APPLICANT: Wellens, Thomas E.
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
: TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe Martens Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: California
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/487,826B
: FILING DATE: 10-SEP-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelsen, Ned
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: NIH121.001CPI1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2710 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Plasmodium falciparum
: US-08-487-826B-12

```

Query Match 26.4%; Score 3197.5; DB 2; Length 2710;

Best Local Similarity 30.6%; Pred. No. 1e-203; Matches 834; Conservative 244; Mismatches 569; Indels 1075; Gaps 74;

```

Qy 4 SCGSGGTQDDADKAVLDEFOQKAVDEYHGEAKNVVSLKGSLSIASLIGTAFVYKSMOT 63
Db 14 AAGGDDIEDSAKMHFRIQKNDVYDKVKEAKERKGLQGLRLS-----BAKFEKNESDP 67

```

[illegible]

Dd	1094	AYKKLRBDMWEANRHOVWRMAMKATGIIICPGHPVDYDIIPONLRMMTEVAENYCAKQSOE	1153
Qy	844	-----	843
Dd	1154	YDKLKICADOMSGDGCKTQGVVDGCKCAACDKYKEEIEKMNBDQWKISDXYLVLVQ	1213
Qy	844	-----	843
Dd	1214	AKTTSTNPGRTVLGDDPDYQQWVDELTPILKASIAARLVKRAAGSPTEIAAARPIPY	1273
Qy	844	-----	843
Dd	1274	STAGYIHOEIGIGCOEJOIFCEKKHGATSTSTYENKEEYFPKOPPEYATACDINRS	1333
Qy	844	-----	850
Dd	1334	QTEPRKKKEENVESAKTIVELIEGKNNGRTVYECNPKBSYPDMDCKNIDISHDGACP	1393
Qy	851	-----	863
Dd	1394	PRROKLLYYIAHESOTENIKRTDNLKDAFIKTAAELEPLSMOYYKSI NDSEAKILDRGL	1453
Qy	864	-----	875
Dd	1454	IPSOFLASMMATTFEDYDIDLTNLNIDISKONDAKAKDKIGKFFSKDG:KSPESGLROEWM	1513
Qy	876	-----	879
Dd	1514	KTNGPEIWKGMCLATKYVTDTONKRIKNDYSYKRVNQSQNPSLIEFAKPOFLRM	1573
Qy	880	-----	879
Dd	1574	IEMGEFCABEROKKENTIKDACEINSTOCCNDAKHRCNOACRAYOE:VENKKKEFSOT	1633
Qy	880	-----	879
Dd	1634	NNFLKANNVQPODEYGYGYKQGVQPIQGENVLQKONNCSGMD:NVLSVSPKEKF	1693
Qy	880	-----	903
Dd	1694	GKYAHKYPEKDCQYOGKHVSIPEPPPPVQPOPEAPTYVVDCSIY-KYLFKDTNNFSDA	1752
Qy	904	CSLYVGNGNSHLGRVCY--TPSGEPTT--SGDKGATCVPPRRRLYIKKIVYMAKRT	957
Dd	1753	GLVY-GKTAPSSKWCIPSPOTKSGAGATTGKSSDSGSJCIPIRRRRLLVGLQDEMAIAL	1811
Qy	958	ESPOASGSEASSTGSGTTPDPSKEALIKAFVESAALETFFLHMRYKEEKKAAVQEGAGHG	1017
Dd	1812	--PQEGAPSHSHA-----DCLRNFQISAALETFFLMDRKKKEKKPPG-DGSOQA	1866
Qy	1018	LPRVEE-GSPEDVPEDK-LKEGKIPDGLRQMFYLYGDIROLLESGSDNTSVSKDPS	1074
Dd	1861	LSOLTSTYSDEEBPPDKLLQNGKIPDPDEFRLMFYLYGDIRLIVHGN--TSDGSGNTNG	1918
Qy	1075	SSNONLKNIVYLASGTEQEREMKNRYKET-KNFKSGTSPASAPLVSHPTQWENNKG	1132
Dd	1919	SNNN--NIVLEASGNKEDMOKIOEKIBOILPNNGSTPLVPRSS--AOTPDKMNNEHAE	1972
Qy	1133	YIWHGMYCALT---SKDKIANKGVKKPOKIEPNEMJAMDE---AKK-----KP	1175
Dd	1973	SIMGMICALYTFEKKNFDTSARQDE--NKIEKDEVEYERKFGSALKNIGASTPYTGYY	2029
Qy	1176	POYOTYNAKLDENSGTSPRTTQTOASSDNTPTTLHFVKRPPTYFRMYIEWGHSFCREKK	1235
Dd	2030	TOYUYEKVKLEDSG---AKTPSASDTP-LISDFVLRPPEFRYLIEWGONFCFKSKH	2083
Qy	1236	RLKOIKVDCKEVENDVG-----RCSGGEGACDSISTHDSYTP:FNCSPPCGKHCSSY	1287
Dd	2084	KLAQIKHECKYEEENGSGSRGGITTRQYSGDGEACNMLPKNDGYPI:LEKPSCAKPCSY	2143
Qy	1288	RKMIERKKIEFHKOSNAYGOOKTDATRRNGNTPFDEKFCCTLETFWPI:AKFLERLKNGBCK	1347



Db 2144 RKMIESGKEFEKOEKAEYDOK -DKVNGSKKNDNGFCETLITTSKAKDLTKL--GPCK 2200  
QY 1348 TNKEYGDDIDIDPEKDSKTFQHTCYCGPCPKFKTNCQNGCVSLNGCNDCKSIDAKEI 1407  
Db 2201 PNNVEGKTIFD--DDKTFKTKDCDCLKFSVCKKDECD-NSKJGDCRKNKSIDATDI 2256  
QY 1408 AKMBSSTDDVMVRSVSDNTFESGDLKDACQAHNIFKGIKRDVWKGCGVYDCEQTN 1467  
Db 2257 ENGVDSTV-LEMRYVSADSKSGFNGDGLFNACRGAGIEGIRKDDMKCRNVCYVCKPEN 2315  
QY 1468 INERTDKEYIQRALKRWRWENFLEDYNNKINDKISHCIKKEGSKCNGECSKCLEK 1527  
Db 2316 VNGAKGHHIIQIRALVKRWYEFEDYNTKIKHISHRIKNGEISPC1---KN--CWEK 2369  
QY 1528 WIEKIAEWENIKRRFNDQYENKQDPYVKSILEELPKIAVVDODNVIKLCVFENSK 1587  
Db 2370 WVDKREKMEKITERFDQYKNDSDDDNVASFLETILPQTDANAKNKYIKLSKFGNSC 2429  
QY 1588 GCTLIISNQ--NKNENDIDMKLKGAKNCPK--PSEKQSDCKEPPPLPDE-- 1639  
Db 2430 GCSASANEQNNNGEYKDAIDCMLKKDKIGECEKHHQTSDECSPTPQPTLEDETLD 2489  
QY 1640 ---DDQNPENTLEPPKFCPT--TORPEKGGETCGNKKEKDESEEPKESGP 1694  
Db 2490 DDIETEAKKNM--PKICENVILKTAQOEDEG--C-----VPAENSEEPAATDSGK 2537  
QY 1695 AAEPAATAESEETETNPPEPGTGAAPSTPA--PRPDTPPLRPQADPEPDSFIL 1751  
Db 2538 ETPEQOTVLRPEEAVPEPP-----PQEKAPAPLPQPTPPTPOLDNHVLTA 2591  
QY 1752 QT-TIPGVALAGSIAFLFK 1772  
Db 2592 VTSTLANSVGIGFATFYFLK 2613

RESULT 5  
US-08-568-459A-10  
; Sequence 10, Application US/08568459A  
; Patent No. 5849306  
; GENERAL INFORMATION:  
; APPLICANT: Stim, Kim L.  
; APPLICANT: Chitnals, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Xin-zhaun  
; APPLICANT: Wellens, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knodde Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: US  
; Zip: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/568,459A  
; FILING DATE: 07-DEC-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israel, Ned  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: NIH121, 001CPI  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 700 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Plasmodium falciparum  
US-08-568-459A-10

Query Match 11.4%; Score 1385; DB 2; Length 700;  
Best Local Similarity 40.8%; Pred. No. 5,5e-84;  
Matches 328; Conservative 106; Mismatches 207; Indels 162; Gaps 34;

Db 113 CAPRRRLHLCNKPNNMNSDSSAKHDLAEVMAKYGESIKTYIPKYDSKPSDF 172  
10 CAPRRRLHLCDY---NSESIDTSTHKLLEVCMAKYGESINTHYTOHQTNEEDSAS 66  
QY 173 PMCTMLARSPADIGDIIRGRDLYLG--NKKKKONGKETEERKLEOKLKEIFPKIH-DNLK 229  
67 QLTVALARSPADIGDIYRGKDLVLYGNKKEQ-----RKLLEOKLDFKTKHAKVYMK 120  
Db 230 DKEAKRYNGD-EDNPFYKLRDMWTANRETWGMATCSKELDSSYFRATCNDTGGPS 288  
121 TNGAQERYIDAKAGDFQRLREDWMTSNRETWYKALICHAPKEANYFTKTAQN-VGKG-- 177  
QY 289 QTHNKKCRDKNKGNAAKPRAGDVTITVPIFYVYQYLRMEFMAEDCRKKKKLEN 348  
178 -TNGQCHC-----TGGD-----VPIFYDYQYLRMEFMAEDCRKKKKLEN 220  
QY 349 LEKCRGKDKSDERYCYSRNGYDEOTISRKGVRMGKGTDCFFACGSYENWIDNORK 408  
221 LOKOCRDYEDN--LYSSGNQYCTKTIYKKGKLVIEHCTNGCSWCMETWIDNCKE 277  
Db 409 FDKOK-KYTRKESIDGG-----RKKRAVGTTKYE--GYEKSEYKLNQDGYTDAFLG 460  
278 FLKOKRKYETEISGSGSGSKSPKRTKRAARSSSSSDNGYESKFKKLKEVYQDQVDFLK 337  
QY 461 LLNNEKACKOTDQGNINFEVNSGGVGVGSGTSGASTNDENKGTYSRSYCOQPC 520  
338 ILNKEGICQKQPOYV--NEKADN-----VDFTNKLYVKTSSREICEPCP 380  
QY 521 DCGVOHKGQWMERKTYKKKRMBSKLYKPIKNGVLLLSKLVKDKMILKKMKKECLT 580  
381 WGLIE-KGPRPW--KVAGDKTCGSAKTKYDPRKNTIDIPVLPRKSOONILKAKKNC-- 435  
QY 581 QNSSDGSVSVYTTGASGNSSEKKELYDEMKCYKHNEVQKVNVOGEVEEDDELKAGGL 640  
436 -----EKGAPGGQIIRK-----WQCYV-----DEHR----- 456  
QY 641 CILPDKKKNKNEVSAK-----SQNNHADIQKTFHDFYVYVWAMHLDKSIHRTKRLKSI- 695  
457 ---PSSKNNNVCYGTWDKFTQCKQT--VKSYNFEFMDWMDHLDHSDVEKTE--LSKGIN 510  
QY 696 --SDGKTMKCRNGCNKKKDCFEKAVKOKETEMKPIKDFYTOEG1-----PRGYU 743  
511 NNTNGNTRCNNNKCKTCGCGFOKVKVEKKQDEMAIKHFKQNDIYQOKGLIYSPYG-- 568  
QY 744 FTYLELILK---LQLEKDETEENTENSLEDAEAEELKHLQKILKENNNLAVNAGTE 799  
569 --VLDVLKGNLLQNIK-LVNHGDT-----DDIKHITKL--LDEEDAVAVVLGSKD 614  
QY 800 QKTLMDKLHNLHDLATKCK---DCPLPREDSRSGSAPSPDIFLRP-----E 846  
615 NTYI-DKLQHEKEQAOCKQKOECEKKAQOESRGSALFREDERTQPADSAGEVEE 673  
QY 847 EKEDENEDDEDEVRDEETAK 869  
674 EDDDDYDEDEDDDDYVDVAVSE 696

RESULT 6  
US-08-487-8268-10  
Sequence 10, Application US/084878268  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,8268  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 700 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-08-487-8268-10

Query Match 11.4%, Score 1385, DB 2, Length 700;  
Best Local Similarity 40.8%, Pred. No. 5.5e-84;  
Matches 328; Conservative 106; Mismatches 207; Indels 162; Gaps 34;

QY 113 CAPERLLCKNKNFNMNSNDSSKAKHDLIAVCAAKYEGSITHTPKYNSKTPGSPF 172  
DB 10 CAPYRRLHCDV--NLSIDTSTTHKLLEVCMAAYEGNSIMTHYQHTORTNEDSS 66  
QY 173 PWCTMLARSFADIGDIIINGRDLYLG--NKKRKONKETEREKLEOKLKEIFKIH-DNLK 229  
DB 67 QLCYVLASFADIGDIVAGKDLVGYDNKKEQ-----RKLLEOKLIDFIKKIKHDVAK 120  
QY 230 DKEAKRYNGD-EDPNFYKLRDWMVTANRETVGAMTCSKELDNSSYFRATCNDTGQGPS 288  
DB 121 TNGAERYIDAKGDFQLREDWMTSNRETVWAKALICHAPKEANYFIKTACN-VGKG-- 177  
QY 289 OTHNRCRCDKDKGANAGRPKAGDGVTVTPYTFDVPVPOYLRFEEWAEDEFCRKKKKLEN 348  
DB 178 -TNGCCHC-----IGGD-----VPTTFDVPVPOYLRFEEWAEDEFCRKKKKLEN 220  
QY 349 LEKQCRGKDKSDERYCGRNGYCEQOTISRKRGVMGKGCQDPCFACGSYEWINDORNO 408  
DB 221 LQKQGRDVEQN---LYCGNGIDCTKITIYKKKGLVIGEHCTNCSYWCNRYETWINDOKKE 277

QY 409 FDKOK-KYTKETSDGCG-----RKRRAVGCTTKYE--GYEKSFYEKLHDGCTVDAPLIG 460  
DB 278 FLKORRYETEISGGSGSKSPKRRRAARSSSSSDNGYESKFRKLTVGYQDDVDFLK 337  
QY 461 LLNNKACKDITDGGKINPEKVENSGGVVGGSGGTSAGSANDENKGFYKSEXCOPCP 520  
DB 338 IINKEGICOKOPVG--NEKADN-----VDFNEXRYKTFSTETCEPCP 380  
QY 521 DGVQHGKGNOMERTKYKRMRSKLYRPINGKAVLLKSLKLVVDDMLLKRMKEPCLT 580  
DB 381 WGLG-KGSPM--KVGDKTGCSAKTKTYDPKNTDIPVLVLPDRSQYLKKRYNFC-- 435  
QY 581 QNSSGSVGSVYTTGASGNSSEKELIDEMKCYKANEVQKVVQGEVE-SDDELGAGGL 640  
DB 436 -----EKGAPGGQIRK-----WQCY-----DEHR----- 456  
QY 641 CILPNPKNKEVSEAK-----SONNHADIOKTFHDPFYVVAHMLKDSIIMPTKRLKSCI- 695  
DB 457 ---PSSKNNKNCVEGTWMDKFTGKOT--VKSYNVEFWMDVHMLDSDVEMKTE-LSKCI 510  
QY 696 --SDGTYKCRNGCNKKKDCFEKRVYKQRETEWKPIDHEKTQEGT-----PEGY 743  
DB 511 NNTNGTCRNNKCKCTDCGCFQKWEKKQOEWMAIKDHRGKQTDIVQKGLVFSPYG-- 568  
QY 744 FTTLELIILK----LQFLKEDTEENTENSLDAEAELEKHLQIILKLENENNLAVYVNACTE 799  
DB 569 --VLDLVYLGKGNLQNIK-DVHGDT-----DDIKIKKL--LDEEDAVAVVLGGKD 614  
QY 800 QKTLMDKLLNHELDNATCK--DCPLPEEDSKSGRSADPSDPIFIRP-----E 846  
DB 615 NNTI-DKLIQHEKEQAEQCKQKQECCEKKAQOESGRSAETREDERTQ2PADSAGEVEE 673  
QY 847 EKEDDEDDEDEVYRDEETAK 869  
DB 674 EDDDDYDEDEDVYQVDVASE 696

RESULT 7  
US-08-568-459A-8  
Sequence 8, Application US/08568459A  
Patent No. 5849306  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/568,459A  
FILING DATE: 07-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CP1  
TELECOMMUNICATION INFORMATION:



```

0Y 1452 WKQGVGVYDICEQTMINER--TDGEYEQIQAILEKRRWENTLEBYNKINDKISHCIRKG 1509
Db 129 WKCAKVCGLDVCGLKNGNSIDKDOQOIIIIHALLKRWEEFLJEDYKINKAKISHTCKRD 188
0Y 1510 EGSKCJINGCEKNSKCLEKWEKIAEWENIKRRFNDQYENKDOOPYNKVLSIEELIPKIA 1569
Db 189 NESTCTNDCKNCTCYEBEMINOKRTMKIKKHKTQENGN---NMKSLVTLIDGALQ 245
0Y 1570 VANDODNVIKLC---VFENSKGCTLLISNTO--NNKENDAIIDCMUKILGVAKNCPGRPS 1623
Db 246 POSDVNKAIFPCSGLTAPEFSGCLNGANDSEKKEGEDYDVLVCMKLNLEKQIOECK-KKH 304
0Y 1624 GER--QSDCKEPPPLD---EEBQNEENTLEPPKFCPTPTOPPEEKGCIGCKNKEEKD 1678
Db 305 GELSVENGSGSCTPLNNTLLEEBPIEENQVAPNCP-----KQYED 348
0Y 1679 EKKESEEPKKEBSGPAEEBAPTAESEETENFPEPPTGPAAPSPAPPTPTPP- 1737
Db 349 KKKEEBETC-----TPASEVPKPPVH 371
0Y 1738 -----LRQADEPDPSTILOTTIPFGVALAGSIAFLFKKTKRASVGNLFQLOIPK 1790
Db 372 VARWRPTPEVEFKIMGRNKKTCYI-----VAEMLKDNKGRJTVGECYR-----K 418
0Y 1791 SDDIDPLKSS---NRIPIYSDRKKGTYIYME---GDSDEKYYAPMSDITDVT--- 1839
Db 419 EYSEWVCDESKIRMOHGACIAPRROKOCCLHYLEKIMINTELKYYAFIKCAAAETFLW 478
0Y 1840 -----SESSEYELD--INDIYVPGSPKYKTLIEVLEPSSGNNTTASAKNTPSDFRNDIQ 1892
Db 479 QNKKKDKNGNAEJDLKGLGGIIPEDFKQMFY-----TPADRIDICL 521

```

```

1 STREET: 620 Newport Center Drive 16th Floor
2 City: Newport Beach
3 STATE: California
4 COUNTRY: US
5 ZIP: 92660
6
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: Floppy disk
9 COMPUTER: IBM PC compatible
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11 SOFTWARE: Patent Release #1.0, Version #1.25
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/568,459A
14 FILING DATE: 07-DEC-1995
15 CLASSIFICATION: 435
16
17 ATTORNEY/AGENT INFORMATION:
18 NAME: Israelson, Ned
19 REGISTRATION NUMBER: 29,655
20 REFERENCE/DOCKET NUMBER: NIH121.001CP1
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE: (619) 235-8550
23 TELEFAX: (619) 235-0176
24 INFORMATION FOR SEQ ID NO: 4:
25 SEQUENCE CHARACTERISTICS:
26 LENGTH: 1435 amino acids
27 TYPE: amino acid
28 STRANDEDNESS: single
29 TOPOLOGY: linear
30 MOLECULE TYPE: protein
31 HYPOTHETICAL: NO
32 ORIGINAL SOURCE:
33 ORGANISM: Plasmodium falciparum
34
35 US-08-568-459A-4

```

```

: RESULT 9
: US-08-568-459A-4
: Sequence 4, Application US/08568459A
: Patent No. 5849306
:
: GENERAL INFORMATION:
:
: APPLICANT: Sim, Kim L.
:
: APPLICANT: Chitnis, Chetan
:
: APPLICANT: Miller, Louis H.
:
: APPLICANT: Peterson, David S.
:
: APPLICANT: Su, Xin-zhaun
:
: APPLICANT: Wellem, Thomas E.
:
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
:
: TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
:
: NUMBER OF SEQUENCES: 37
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Knodde Martens Olsson & Bear

```

[illegible]

```

QY 835 DPEIDIFPREEKEDDENEDDEDEVRDEDEAKETTESGATDTTSLDVCPIYGVKVL 894
Db 463 -----DKN-----SVPTNKV-----473
QY 895 KDNESLADACSLKYGNNRLGMRCVPSGPTTSSDKNGAICVPRRR-----RLY 946
Db 474 -----WEC-----KNPYLSTKD--VCVPRRQELCLGNIDRIY 505
QY 947 IKKIVDAKTESPQASGEASSTSGSTTPPDSKELLKAVEBAAT-ETFFLHRYKEE 1005
Db 506 DKNLL-----MIKEHILALAIYESHLLKRYKKNK 534
QY 1006 KKAQAEGAGHGLPRVEGSEPEYDEPKLKEGKIPDGLRQMFYTLGDRILFGSNDT 1065
Db 535 -----DOK-----EVCKITNTFADIRIIT--GGTDY 559
QY 1066 TYSKIDTPSSNDNLKNIIVLASGSTEOERKMKYKEIKFRKCSSTERSAPNLVSHPT 1125
Db 560 W-----NDLSNRKLLVGKINTNSKYVHRNK-KNDKLFK-----DE 592
QY 1126 WMENNGKXIMHGMVCAKLSKDKIAKVEKKRQKTEPENLMDANKKPKPQYQYTNKL 1185
Db 593 WMAKIKKDVWN--VISWVFKDVT--CKEDTEN-----622
QY 1186 DENSGTSPRTTQTOASSDNTPTTLHFVKRPTYPFMEWGESFCRERKKRLKOIKYDK 1245
Db 623 -----IPQFRFSEMDDDYCODKTKMIEFLKVECK 653
QY 1246 VENGDCVRCSGDGACDSISTHDYSTVPSFNCPCGKHCSRYRWIEKRTIEPKOSNAY 1305
Db 654 EK-----PCBED-----NCKSKNSKEMISKKEEYKNOAKOY 687
QY 1306 GOKTADATRNNGNTFDEKFCLETPWDAKFLRLKKNPCPKTKVEYGGDDIDENKSKT 1365
Db 688 QEYO-----KGNNTMYSEF-KSIRPEVYLKYSKCSN-----LNEDEKKE 729
QY 1366 FOHTEY--CGCPKFR-----TNCQNGNCVSGLNG-----CDEKSIDA 1404
Db 730 ELHSDYKKNKCMCEYKDVPIISIRNNQTSQEAVPENPTIAHRTPTISSEPKMGEO 789
QY 1405 KEIAKMSSTTDVYMRVSDNDNTFEGDDLKDACOHANIEKIRKDY--WKGVCYGVDI 1462
Db 790 KE-----RDDDSLKTSISVSPENSRPELDAKT--SNLTK-LKGDVDISMRKAVIGSSP 839
QY 1463 CEQTNINERTDGKEYIQRALFRKRVENPLEDYKINDKISHCKKGGSGKINGCEKNS 1522
Db 840 NDNINATEOGDN-----ISGV--NS 857
QY 1523 KCLEKMTIEKIAEWENIKRRENDQYENKDDPYNVKSILEELIPKIAVNDQDVNIKLV 1582
Db 858 KPLSDVDRP-----KKELEDO--NSDESEETVW--HISKSPINNOD-----897
QY 1583 FENSKGCTLSINTONKNENDAI-----CMLKKLVAKKNCPRKSGEKOSDCKEPPP 1635
Db 898 -DSGSGSATYSESSSSNTGLSIDDRNGDTFVRTQDANTEDVIRKERADKDEDEKG--953
QY 1636 LPDEEDONPEENTLEPPKFCPTQPREK-----GGTCGNKEEKKDEKKESEERA-1668
Db 954 -ADERHSTSE-----SLSSPEEKMLJNEGNSLNHBEVKEHTSNSNOVQSG-1001
QY 1689 -----KEESGPAABEPAPTAASEETETNPEPPGTGPAAPSTPAPPTDTPPLRPQ 1741
Db 1002 GIYVANNVEKELKDTLENPSSLDGKAHELESPNLSSDOMSTPG-----1048
QY 1742 ADEPFDSSTIIQTTIPFVALALGSIAPFLKKTAKVAGNLFQILQIPKSDYDIPTLKSS 1801
Db 1049 -----PLDNTSEETTERI-----SN 1063
QY 1802 NRYIPYVSDRKKTYYIMEGDSDDEKAYAFMSDTTDTYSSSEVEELDINDIYPPGSPKY 1861
Db 1064 NEY--KVNER-----EDERTYLTKEVEDIVLASHMNRSDDCGLY-----1100

```

```

QY 1862 KTLIEVLEPSGNNTTASGKNTPSDTRNDIONDGISSKITDNEPMOLKKEPISNMLQ 1921
Db 1101 -----DENSDLSLVNDESEDA--EAKMKGND-----TSEMSHNS 1132
QY 1922 PNDVPNDYTSNGSNSTNTTITTSRHNVDNNTTMSRDNMEENILPISIHGNIYSGEY 1961
Db 1133 SQHIESDOQKNDKMTVGLGCTT--HVONEISVPTGEIDEK--LRESKSKYHKAEE 1186
QY 1982 SYNVMNMSNDIPINRDNVYSGIDLINDLSGCKPIDIYDEVLKREKNEPCTENTKR 2041
Db 1187 RLSHTFDIKIN--PEDRSNTHLHKDIRNE-----ENERH 1219
QY 2042 TSTQNV-----AKTNSDPINHOELFHKWLDRRHDCERKKNKEDILNK 2086
Db 1220 LTNQININISQERDLQKHGFHTMNNLHGDGYERSQIHN--SHHGNNQDRGNSGVNLM 1276
QY 2087 LKEEMKENINNSGKTTNSDNKPSHNVLNTVSIQIDMDNPKTKNETTMDTQDXSTM 2146
Db 1277 RS--NNNNFNINPSRYNLKYD-----KLDLDYENRNDSTTKELIKLAEI 1320
QY 2147 DTILDDLE-KYND 2158
Db 1321 NKCENEISVKICD 1333

```

```

RESULT 10
US-08-487-826B-4
; Sequence 4' Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Slim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; Zip: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487, 826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-0176
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-487-826B-4

```



```

:
: APPLICANT: Miller, Louis H.
: APPLICANT: Peterson, David S.
: APPLICANT: Su, Xin-zhaun
: APPLICANT: Wellens, Thomas E.
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
: TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe Martens Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: California
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/568,459A
: FILING DATE: 07-DEC-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelsen, Ned
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: NIH121.001CPI1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 362 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: Internal
: ORIGINAL SOURCE:
:
: US-08-568-459A-18

```

```

Query Match          3.1%; Score 379.5; DB 2; Length 362;
Best Local Similarity 31.3%; Pred. No. 1,7e-17;
Matches 93; Conservative 14; Mismatches 149; Indels 41; Gaps 5;

```

```

OY 113 CAPPRRLHLCKNFPNMNSNDSSAKHDLAEVCMAYEGESIKTHYPKYSKYPGSDF 172
    ||:|||||:|:|:|
    2 CAPYRLHLCDY---NLKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 58
OY 173 PNCMTLARSFADIDIGDIRGRDLYLG--NKKKKQNGKETEREKLEOKLEIFKKIHDNLKD 230
    ||:|||||:|:|:|
    59 QLCIVLARSFADIDIGDIRGRDLYLGDKKXXXXXXXXXXXXXXXXXXXXXXXXXXXX 118
OY 231 KEAOKRYNGDEDPNFYKLREDMWTANRETVGAMTCSKELDNSSYFRATCNDTGOGPSQT 290
    ||:|||||:|:|:|
    119 XXXXXXXXGDD---FQOLREDMWTNSRETVMKALICHAXXXXXXXXXXXXXXC----- 164
OY 291 HNKCRDCKDKANAGKPRAGGDVTVPTFYDVPQYLRFMEEWAEDFCRRKKKKLENL 350
    ||:|||||:|:|:|
    165 -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXVPQYLRFMEEWAEDFCRRKKKKLENLQ 214
OY 351 KQCRGKDKSDERYRCRSRNGYDCEOTISRKGVRMGKGTDCFFACGSEYENWIDNQRK 407
    ||:|||||:|:|:|
    215 KQ-----CXKXXXXXXXXXXXXXXXXXXXXCTNCNSVCMRYETWIDNQRK 259

```

```

RESULT 12
US-08-487-826B-30
; Sequence 30, Application US/08487826B
; Patent No. 5993827

```

```

:
: GENERAL INFORMATION:
: APPLICANT: Sim, Kim L.
: APPLICANT: Chitnis, Chetan
: APPLICANT: Miller, Louis H.
: APPLICANT: Peterson, David S.
: APPLICANT: Su, Xin-zhaun
: APPLICANT: Wellens, Thomas E.
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
: TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe Martens Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: California
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/487,826B
: FILING DATE: 10-SEP-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelsen, Ned
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: NIH121.001CPI1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 30:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 362 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: Internal
: ORIGINAL SOURCE:
:
: US-08-487-826B-30

```

```

Query Match          3.1%; Score 379.5; DB 2; Length 362;
Best Local Similarity 31.3%; Pred. No. 1,7e-17;
Matches 93; Conservative 14; Mismatches 149; Indels 41; Gaps 5;

```

```

OY 113 CAPPRRLHLCKNFPNMNSNDSSAKHDLAEVCMAYEGESIKTHYPKYSKYPGSDF 172
    ||:|||||:|:|:|
    2 CAPYRLHLCDY---NLKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 58
OY 173 PNCMTLARSFADIDIGDIRGRDLYLG--NKKKKQNGKETEREKLEOKLEIFKKIHDNLKD 230
    ||:|||||:|:|:|
    59 QLCIVLARSFADIDIGDIRGRDLYLGDKKXXXXXXXXXXXXXXXXXXXXXXXXXXXX 118
OY 231 KEAOKRYNGDEDPNFYKLREDMWTANRETVGAMTCSKELDNSSYFRATCNDTGOGPSQT 290
    ||:|||||:|:|:|
    119 XXXXXXXXGDD---FQOLREDMWTNSRETVMKALICHAXXXXXXXXXXXXXXC----- 164
OY 291 HNKCRDCKDKANAGKPRAGGDVTVPTFYDVPQYLRFMEEWAEDFCRRKKKKLENL 350
    ||:|||||:|:|:|
    165 -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXVPQYLRFMEEWAEDFCRRKKKKLENLQ 214
OY 351 KQCRGKDKSDERYRCRSRNGYDCEOTISRKGVRMGKGTDCFFACGSEYENWIDNQRK 407
    ||:|||||:|:|:|
    215 KQ-----CXKXXXXXXXXXXXXXXXXXXXXCTNCNSVCMRYETWIDNQRK 259

```

```

RESULT 13

```





Oy 353 C-RGDKSDEYRCSRNGIDCEQTISRKKGYMGKGCLDCCFACSGSYENWTDNRK 407  
| | : || | :||::  
Db 242 CXXXXXXXXXXXXXXXCCCCCSCSLCYACPVDWTINQKE 297

RESULT 15  
US-08-568-459A-6  
; Sequence 6, Application US/08568459A  
Date: 08/08/2008

; GENERAL INFORMATION:

1  
 2 APPLICANT: Sim, Kim L.  
 3 APPLICANT: Chitnis, Chetan  
 4 APPLICANT: Miller, Louis H.  
 5 APPLICANT: Peterson, David S.  
 6 APPLICANT: Su, Xin-zhaun  
 7 APPLICANT: Wellens, Thomas E.  
 8 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
 9 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE  
 10 NUMBER OF SEQUENCES: 37  
 11 CORRESPONDENCE ADDRESS:

Db	202	IGDIVRGIDVW-----	RDINTNKLKSKPCKIF-----	MGGSNKKKKQNDNNE--	24.3	
Qy	245	FYKLEDMWTANRETWVGAMTCSKELDNSSYFRATCNDTGGPSQTHNKKCRCDKDGANA		30.4		
Db	244	-----RNMKEKQRMILIMSSW--	-KHIEPKG-----	KRHNN-----	27.4	
Qy	305	GKPKAGODVATVTPYPPVYQVLYRMEPEANEDFCRRKKKKKLEVLKQCKGDKSDERY		36.4		
Db	275	-----	-FEKIPQFLRMKLKWDDEFCCEKGTGVKOLEIKENKNSK-----	31.5		
Qy	365	CSRNGYDEQOTISRKHGKVRMGKGTDFEFACGSEYENNIDN-----	ORKOFDKOKKVTX	41.7		
Db	316	-----	KCKMCCSYSEKNIKKEKNENYNIQSKKFPDSKFLNK	35.0		
Qy	418	EISDGGKHKRAVGTTKTYEGYEKSPFYBKLNKGITGVADFLGLNNKCAKCIDTGGCT		47.7		
Db	351	-----	-KNLNYNKE-----	DSKAYLRSESKQCSNTE-----	37.6	
Qy	478	NFKEVNCGGVVGGSGGTSGASGTINBNKGT F-YRSEYQOPCPDGCQVHGKGQWOMERT		53.6		
Db	377	-----	FND-----	TTFEPKKYKEACWY-----	35.5	
Qy	537	KVKKRWSKLYKPINGKAVLLLSLYVKDMILKMKKEFCLTQNSSDGSVGYTTGA		56.6		
Db	396	-EWPSSSKALKALPIKTNVPIEESK--	SELSLTDKSKN--	TPNSS-----	43.6	
Qy	597	SGGSEKKEK-----	YDEKWKYK--	HNEYOKVNVQGEV-----	EDDELKGA	63.7
Db	437	GGGVGYGRQJIKRDDVHHDGPEKVEKSEKEVPKIDAIVKTENEFTSNRNDIEGKESKD		49.6		
Qy	638	GGCLILPNPKNKRYVSEAKSONNIADLOKTFHDFEYVVAHMLKDSI--	HWRPKRLKSC	69.4		
Db	497	HSSEVYHSDKIDNEBPQRYVSE-NLPKIEEKES-----	SDSLPIRTHIEAKQSS	54.5		
Qy	695	ISDKETMKRCNGCNKKCDCEFEKNVQKQETEMPKIKDHFKTQEGIEPESYFTLELILKQ		75.4		
Db	546	NSSDNDAVVAVSGRESKDVNLHTSRIKENE-----	EGV-----	IKTDSSKSIE	58.9	
Qy	755	FLKEDTEENTENSIDAEBAEELKHLQILKLENNNLAAYNAGTEQKTLMDKLNLHEND		81.4		
Db	590	ISKIPSDQNNHSDLSQANABD-----	SNOGNETINPSTENKL--	KEIHYTSD	63.7	
Qy	815	A-----TCKDCDPLPEE-----	DK-----	SRCGRADSPSDIRIPPEEKEDENE	85.4	
Db	638	SDDHGSKIKSIEPEKELTEBEPRLDTKTESAIAQDKHHEYSKASDIT--	QSEIHNSNRD	63.5		
Qy	855	DDDEDEVADEETAKETEGESATDT--	TSLDVCPITVGVLTKDNESLQDACSILKYGN	91.1		
Db	696	RIVSESVQDSSGSSMSTESTIRTDNKKDKFTSEDIAPSI-----	NGR	73.6		
Qy	912	NSRL	915			
Db	737	NSRV	740			

Search completed: May 1, 2001, 13:04:37  
Job time: 150 sec

---

**THIS PAGE BLANK (USPTO)**